



Clonage et caractérisation de deux gènes codant des enzymes lipolytiques de la microalgue *Isochrysis galbana*

Vincent Kerviel

► To cite this version:

Vincent Kerviel. Clonage et caractérisation de deux gènes codant des enzymes lipolytiques de la microalgue *Isochrysis galbana*. Sciences agricoles. Université du Maine, 2014. Français. NNT : 2014LEMA1017 . tel-01149642

HAL Id: tel-01149642

<https://theses.hal.science/tel-01149642>

Submitted on 7 May 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Thèse de Doctorat

Vincent KERVIEL

Mémoire présenté en vue de l'obtention du grade de Docteur de l'Université du Maine

sous le label de l'Université Nantes Angers Le Mans

Ecole doctorale : V.E.N.A.M.

Discipline : Sciences du Vivant

Spécialité : Biologie des Organismes

Unité de recherche : EA 2160 Mer Molécules Santé (MMS)

Présentée et soutenue publiquement le 23/09/2014

Thèse N° :

Clonage et caractérisation de deux gènes codant des enzymes lipolytiques de la microalgue *Isochrysis galbana*

JURY

Rapporteurs :

Dr. David COBESSI, Chargé de Recherche HDR-CNRS, IBS-Grenoble

Dr. Frédéric DOMERGUE, Chargé de Recherche HDR-CNRS, Université de Bordeaux

Examineurs :

Pr. Benoît SCHOEFS, Professeur des Universités, Université du Maine-Le Mans

Pr. Elisabeth PETIT-TEIXEIRA, Professeur des Universités, Université d'Evry-Val d'Essonne

Directeur de Thèse :

Pr. Françoise ERGAN, Professeur des Universités, Université du Maine-Laval

Co-encadrant de Thèse :

Dr. Céline LOISEAU, Maître de Conférences, Université du Maine-Laval

Membre invité :

Dr. Jean-Luc JUNG, Maître de Conférences, Université de Bretagne Occidentale-Brest

Membre invité

et co-encadrant de la thèse : Dr. Laurent POISSON, Maître de Conférences, Université du Maine-Laval

REMERCIEMENTS

Les travaux présentés dans ce mémoire ont été réalisés au sein du laboratoire Mer, Molécules, Santé (MMS), dans sa composante basée à Laval en Mayenne dans l'équipe ChimiMar de l'axe 3 «CHIM » de MMS. L'équipe prend place au sein du Département de Génie Biologique de l'IUT de Laval.

Je remercie les Collectivités locales Mayennaises pour avoir financé ces travaux de Recherche.

Je remercie Monsieur David Cobessi, Chargé de Recherche HDR au CNRS et Monsieur Frédéric Domergue, Chargé de Recherche HDR au CNRS, qui m'ont fait l'honneur d'accepter de juger ce travail et d'en être rapporteurs. Je remercie Madame Elisabeth Petit-Teixeira, Professeur à l'Université d'Evry-Val d'Essone et Monsieur Benoît Schoefs, Professeur à l'Université du Maine pour avoir accepté de participer à ce jury de thèse.

Je remercie Françoise Ergan, Professeur de l'Université du Maine, pour m'avoir accueilli dans son laboratoire de recherche et pour le soutien qu'elle m'a apporté tout au long de ce projet.

J'exprime toute ma reconnaissance envers Céline Loiseau et Laurent Poisson, Maitres de Conférences à l'Université du Maine, pour avoir co-encadré ce travail.

Céline, merci d'avoir su être là pour tous les moments de cette thèse, les bons, les mauvais et les excellents moments et particulièrement pour le temps consacré à la correction de ce manuscrit.

Je remercie chaleureusement Josiane Hérault, pour avoir toujours su me donner des raisons d'avancer, de croire et de persister dans ce projet. Des premières difficultés jusqu'ici, tu auras toujours été là pour avoir ce regard si précieux sur mon travail, merci.

Je remercie Gaëlle Pencreac'h, pour avoir su être à l'écoute, pour les nombreuses discussions sur la pédagogie de l'enseignement, sur la recherche dans son ensemble et l'aboutissement de ce projet en particulier.

Je remercie l'ensemble des enseignants-chercheurs du laboratoire MMS de Laval, pour leur gentillesse et leurs précieux conseils.

Je tiens à remercier également l'ensemble des enseignants du Département Génie Biologique et particulièrement Françoise Lourdaïs et Gilles Braud pour m'avoir permis d'apprécier le travail pédagogique et d'orientation à réaliser auprès des étudiants. J'attendais tellement de transmettre des connaissances, vous avez su me guider et rendu cette expérience... géniale.

Je remercie enfin le service technique du Département. Isabelle Martin, Frédérique Guéno et Rose-Marie LeRoux qui animent et gèrent l'ensemble des matériels et commandes, avec vous rien n'est impossible.

Je remercie toutes les personnes avec qui j'ai pu partager tout au long de ces quatre années au sein de l'IUT de Laval et du PRES des Pays de la Loire, les pauses café, les repas, l'organisation de congrès, ce sont des moments privilégiés.

Je pense très fort à Catherine, Muriel, François, Marie-Claude pour l'esprit de famille qu'ils m'ont témoigné durant ces 4 ans.

Merci à tous mes amis, thésards, sportifs, docteurs, alsaciens, mayennais, chinois, marocains, post-docs, DR, CR1, CR2, vous êtes loin, je ne vous oublie pas, la vie universitaire ou associative nous a permis de nous connaître, vous êtes ma seconde famille. Merci à Gaël, Nath', Céline, Aude et Guillaume pour votre soutien, merci à Anaïs & Co pour le soutien durant ma rédaction. Merci Jo, Nadège pour votre patience. Merci à Franck, merci à Gisèle et un merci tout particulier à Sophie qui m'a soutenu et a permis que tout cela se réalise.

A ma famille,

A mon Grand Père, là où tu es j'espère que tu es fier de moi.



Equipe Laval
ChimiMar



Laboratoire Mer, Molécules, Santé EA 2160

Equipe ChimiMar – Laval

IUT de Laval – 52 rue des Docteurs Calmette et Guérin

53020 Laval Cedex

RESUME/ABSTRACT

Résumé

Les enzymes lipolytiques sont des ester hydrolases impliquées dans le métabolisme lipidique. Leurs caractéristiques se sont révélées être des atouts dans de nombreuses applications industrielles. Chez les microalgues, l'isolement et la caractérisation de ces enzymes d'un point de vue structural et fonctionnel restent des domaines de recherche peu explorés à ce jour.

Certaines espèces bénéficient pourtant de contenus en lipides intéressants, source de matière première pour les industries de l'agroalimentaire ou de l'énergie. Par exemple, l'acide docosahexaénoïque (DHA), un acide gras polyinsaturé de la série des oméga 3, est reconnu pour ses propriétés en santé humaine. Parmi de nombreuses espèces, *Isochrysis galbana*, une microalgue unicellulaire appartenant à la classe des Prymnesiophycées est considérée comme une source possible de DHA. La présence d'acides gras libres a été montrée par l'analyse des lipides, suggérant la présence d'enzymes lipolytiques potentiellement intéressantes pour leur sélectivité et leur spécificité de substrat.

L'analyse d'une banque de marqueurs de séquences exprimées a permis l'identification de séquences susceptibles de coder des enzymes lipolytiques. Les ARN messagers ont été extraits et les ADN complémentaires ont été clonés.

Ce travail de thèse présente l'analyse et le clonage de deux gènes codant une ester hydrolase putative et une thioestérase putative, issues de la microalgue *Isochrysis galbana*.

Les deux séquences codent des protéines de poids moléculaires de 35,41 kDa et de 42,31 kDa. Elles montrent 30 à 40 % d'identité et de similarité avec des hydrolases α/β notamment des carboxylésérases de différents organismes. Les séquences protéiques ont permis l'identification du pentapeptide consensus Gly-X-Ser-X-Gly caractéristique des enzymes lipolytiques et les acides aminés Ser/Asp/His de la triade catalytique.

Les deux séquences codantes ont été clonées et exprimées dans la levure *Saccharomyces cerevisiae* et la bactérie *Escherichia coli*. Le clonage dans *E. coli* a permis d'identifier à la taille attendue une protéine par Western blot. En présence de cette protéine, la composition en acides gras des lipides de la bactérie a été modifiée. L'analyse CPG a notamment montré une augmentation des proportions en acides gras C16:1 et C18:1 par rapport au témoin. Ce résultat permet de caractériser l'activité thioestérase pour *IgTeCe*.

Mots clés : Ester hydrolase, *Isochrysis galbana*, Expression de protéines, Acides gras.

Abstract

Lipolytic enzymes present in all known species play a key role in lipid metabolism and are involved in several industrial processes. They catalyze lipid hydrolysis and synthesis. Actually and particularly in microalgae, isolation and characterization of this type of enzyme remains an unexplored research area.

The potential of the lipidic content of microalgae in food industry or energy field requires specific lipolytic enzymes. Docosahexaenoic acid (DHA), an $\omega 3$ poly unsaturated fatty acid ($\omega 3$ PUFA) is well known for its beneficial effects on human health. Among many species, *Isochrysis galbana*, a unicellular marine microalga belonging to the Prymnesiophyceae class, is considered as a potential alternative source of DHA.

Lipid analysis of *I. galbana* shows free fatty acids and suggests the presence of lipolytic enzymes with potential interesting selectivities and substrate specificities. Analysis of incomplete expressed sequence tag (EST) listed in the EST bank of *Isochrysis galbana*, identified incomplete genes that encode lipolytic enzymes. Messenger RNAs were extracted, characterized and cloned.

This work describes the analysis and cloning of two genes encoding a putative ester hydrolase and a putative thioesterase in marine microalgae *Isochrysis galbana*. Sequences encode two proteins with predicted molecular weights of approximately 35,41 kDa and 42,31 kDa. Slight similarity and identity (from 30 to 40 %) were observed between the gene sequence and various α/β fold hydrolase found in diverse phyla (including carboxylesterase).

Sequences also included the consensus Gly-X-Ser-X-Gly and the catalytic triad Ser/Asp/His. To characterize the predicted enzymatic functions, an experimental procedure was introduced: coding sequences were cloned into expression vectors and expressed in *Saccharomyces cerevisiae* and in *Escherichia coli*.

Western blot identification of recombinant enzyme shows a convenient protein production in bacteria. Furthermore, the expression of the protein in *E. coli* shifted the fatty acid composition predominantly towards C16:1 and C18:1 fatty acids. The enzyme called *IgTeCe* showed a thioesterase activity.

Keywords : Ester hydrolase, *Isochrysis galbana*, Protein expression, Fatty acids

TRAVAUX RELATIFS A CETTE ETUDE

Les travaux de cette thèse ont fait l'objet d'un article dans un journal scientifique, de trois posters lors de congrès nationaux et d'une présentation orale lors d'un congrès international.

Article

Journal of Applied Phycology : accepté en Janvier 2014

Cloning and expression of a gene from *Isochrysis galbana* modifying fatty acid profiles in *Escherichia coli*.

Kerviel V., Herault J., Dumur J., Ergon F., Poisson L. and Loiseau C.

DOI: 10.1007/s10811-014-0240-7

Présentation orale

Molecular studies on a lipolytic enzyme from the microalga *Isochrysis galbana*

Ergon F., Kerviel V., Héroult J., Poisson L., Loiseau C.

Marine lipids session, 12th Euro Fed Lipid Congress, Montpellier, France, 14-17 septembre 2014.

Posters

Molecular studies on a lipolytic enzyme from the microalga *Isochrysis galbana*.

Kerviel V., Herault J., Poisson L., Ergon F. and Loiseau C.

Annual congress of The French Society of Biochemistry and Molecular Biology, Grenoble, France, 20-23 novembre 2012.

Preliminary studies on lipolytic enzyme from micro alga *Isochrysis galbana*.

Kerviel V., Dumur J., Herault J., Poisson L., Ergon F. and Loiseau C.

Annual congress of the lipid Group of The French Society of Biochemistry and Molecular Biology, Lyon, France, 25-28 octobre 2011.

Over-expression and characterization of putative lipolytic enzymes from newly isolated genes of *Isochrysis galbana*.

Kerviel V., Poisson L., Herault J., Ergon F., and Loiseau C.

Young Scientist Forum of The French Society of Biochemistry and Molecular Biology, Strasbourg, France, 14-17 décembre 2010.

TABLE DES MATIERES

<i>Remerciements</i>	<i>iii</i>
<i>Résumé/Abstract</i>	<i>vii</i>
<i>Travaux relatifs à cette étude</i>	<i>ix</i>
<i>Table des matières</i>	<i>xi</i>
<i>Liste des figures</i>	<i>xvii</i>
<i>Liste des tableaux</i>	<i>xix</i>
<i>Liste des abréviations</i>	<i>xxi</i>
<i>Code des acides aminés</i>	<i>xxii</i>
<i>Introduction</i>	<i>3</i>
<i>Chapitre I : Synthèse bibliographique</i>	<i>9</i>
1. Les enzymes lipolytiques	9
1.1. Définitions	9
1.2. Structure et mécanisme	12
1.2.1. Le système α/β hydrolase	13
1.2.1. Cinétique enzymatique des enzymes lipolytiques	15
1.2.2. Spécificités de substrat des enzymes lipolytiques	17
1.3. Les enzymes lipolytiques à travers le monde vivant	20
1.3.1. Les enzymes lipolytiques végétales	20
1.3.2. Les enzymes lipolytiques animales	23
1.3.3. Les enzymes lipolytiques de micro-organismes	28
1.4. Les enzymes lipolytiques de microalgues	33
1.5. Les applications industrielles des enzymes lipolytiques	36
1.5.1. Méthodes de production des lipases	37
1.5.2. Applications dans les détergents	37
1.5.3. Applications dans l'industrie textile et papier	38
1.5.4. Applications dans l'industrie agroalimentaire et laitière	38
1.5.5. Autres applications	40
1.5.6. La production des monoglycérides	40
1.5.7. Enzymes de microalgues et applications énergétiques	40

2.	Caractériser une activité lipolytique par biochimie et biologie moléculaire	42
2.1.	Préambule	42
2.2.	L'approche biochimique	42
2.2.1.	Méthodes en milieu de culture solide	42
2.2.2.	Méthodes par titration	43
2.2.3.	Méthodes colorimétriques, turbidimétrique et fluorimétrique	44
2.2.4.	Méthodes chromatographiques	45
2.2.5.	Méthodes immunologiques et radioactives	46
2.2.6.	Autres méthodes	46
2.3.	L'approche bioinformatique : gènes, domaines, motifs et séquences consensus des enzymes lipolytiques	47
2.3.1.	Les bases de données	48
2.3.2.	Identification d'enzymes lipolytiques par homologie de séquences	51
2.4.	Identification et caractérisation d'une fonction lipolytique chez <i>Isochrysis galbana</i>	52
2.4.1.	Indices biochimiques	52
2.4.2.	Les gènes d' <i>Isochrysis galbana</i> appartenant au métabolisme lipidique	53
3.	Systèmes hétérologues et expression d'enzymes lipolytiques recombinantes	57
3.1.	Préambule	57
3.2.	La production hétérologue de protéines	57
3.2.1.	Hôte procaryote : <i>Escherichia coli</i>	57
3.2.2.	Hôte procaryote : <i>Bacillus</i> sp.	64
3.2.3.	Hôtes eucaryotes : les levures et les champignons	65
3.2.4.	Hôtes eucaryotes supérieurs	68
3.3.	Les principaux vecteurs d'expression	70
3.4.	Bilan	76
3.5.	L'expression hétérologue d'enzymes lipolytiques	77
	Chapitre II : Matériel et méthodes	87
1.	Matériels biologiques	87
1.1.	<i>Saccharomyces cerevisiae</i>	87
1.2.	<i>Escherichia coli</i>	87
1.3.	Plasmides de clonage et vecteurs d'expression	87
1.3.1.	<i>Saccharomyces cerevisiae</i>	87
1.3.2.	<i>Escherichia coli</i>	88
1.4.	Milieux de culture	89
1.4.1.	Milieu de culture pour <i>Saccharomyces cerevisiae</i>	89
1.4.2.	Milieu de culture pour <i>Escherichia coli</i>	90
1.5.	Conditions de culture et de conservation	90

1.5.1.	Suivi de croissance -----	90
1.5.2.	Méthodes de conservation -----	90
1.5.3.	Méthodes de culture de <i>Saccharomyces cerevisiae</i> -----	91
1.5.4.	Méthodes de culture d' <i>Escherichia coli</i> -----	91
2.	Etude de l'expression chez <i>Saccharomyces cerevisiae</i> -----	92
2.1.	Préambule -----	92
2.2.	Extraction du vecteur <i>pYES2.1</i> contenant les gènes présents dans <i>Saccharomyces cerevisiae</i> -----	92
2.3.	Amplification par PCR -----	93
2.4.	Détermination de la présence d'un insert dans le vecteur <i>pYES2.1</i> -----	94
2.5.	Electrophorèse et purification des produits amplifiés -----	94
2.6.	Quantification des acides nucléiques -----	95
2.7.	Expression d' <i>IgTeCe</i> et <i>IgEst3</i> -----	95
3.	Etude de l'expression chez <i>Escherichia coli</i> -----	96
3.1.	Préambule -----	96
3.2.	Extraction des gènes <i>IgTeCe</i> et <i>IgEst3</i> depuis le plasmide <i>pYES2.1</i> : protocole " K7-extract " -----	96
3.3.	Clonage et ligation -----	96
3.4.	Préparation de cellules d' <i>Escherichia coli</i> et transformation avec le plasmide de clonage. -----	97
3.5.	Extraction de plasmides à partir d' <i>Escherichia coli</i> -----	98
3.5.1.	Mini préparation de plasmides pour la sélection de clones positifs -----	98
3.5.2.	Midi préparation de plasmides pour sous-clonage et séquençage -----	98
3.6.	Validation d'un clone -----	99
3.6.1.	Validation par digestion enzymatique -----	99
3.6.2.	Validation par séquençage -----	99
3.7.	Clonage dans les vecteurs d'expression -----	100
3.7.1.	Amplification et ajout de sites de restriction -----	100
3.7.2.	Digestion enzymatique pour l'insertion dans un plasmide d'expression -----	100
3.7.3.	Ligation et transformation -----	101
3.8.	Validation de l'insertion dans les plasmides d'expression -----	101
3.8.1.	Validation par séquençage -----	101
3.9.	Préparation de cellules d' <i>Escherichia coli</i> BL21 et transformation avec les plasmides d'expression -----	101
3.9.1.	Préparation de cellules électrocompétentes -----	101
3.9.2.	La culture d' <i>Escherichia coli</i> – tests d'expression -----	102
4.	Recherche de transcrits (RT-PCR) -----	103
4.1.	Préparation d'ARN totaux -----	103
4.2.	Obtention de la matrice ADN complémentaire -----	104
4.3.	Evaluation de la RT-PCR -----	104

5.	Techniques de biochimie	105
5.1.	Extraction des protéines de <i>Saccharomyces cerevisiae</i>	105
5.2.	Extraction des protéines d' <i>Escherichia coli</i>	105
5.3.	Dosage de protéines	106
5.4.	Séparation électrophorétique	106
5.4.1.	SDS-PAGE	106
5.4.2.	Coloration des protéines	108
5.5.	Détection immunologique	108
5.5.1.	Dot blot	108
5.5.2.	Western blot	109
5.6.	Chromatographie et purification	109
5.6.1.	Préambule	109
5.6.2.	Tampons utilisés	110
5.6.3.	Procédure de purification	110
6.	Caractérisation des enzymes	111
6.1.	Concentration et dialyse	111
6.2.	Recherche d'activité	111
6.2.1.	Test d'hydrolyse de substrat gélosé	111
6.2.2.	Test d'hydrolyse de substrat chromophore	112
6.3.	Composition des lipides	112
6.3.1.	Extraction des lipides totaux	112
6.3.2.	Saponification et méthylation	113
6.3.3.	Analyse par chromatographie en phase gazeuse des esters méthyliques	113
7.	Méthodes bioinformatiques	114
7.1.	Manipulations et outils de comparaison de séquences	114
7.2.	Profil physico-chimiques des protéines et prédictions de structures	115
7.3.	Analyses et optimisation de prises de vue expérimentales	115
	Chapitre III : Résultats et discussion	119
1.	Préambule	119
2.	Etudes <i>in silico</i> des deux séquences codantes d'<i>Isochrysis galbana</i>	119
2.1.	Etude d' <i>IgEst3</i>	119
2.1.1.	Prédictions des caractéristiques d' <i>IgEst3</i>	121
2.1.2.	Prédictions de la fonction enzymatique par recherche d'homologies	122
2.1.3.	Prédictions de l'organisation structurale	129
2.1.4.	Bilan des analyses <i>in silico</i> pour la séquence codante <i>IgEst3</i>	131

2.2.	Etude d' <i>IgTeCe</i> -----	133
2.2.1.	Prédictions des caractéristiques d' <i>IgTeCe</i> -----	134
2.2.2.	Prédictions de la fonction enzymatique par recherche d'homologies -----	135
2.2.3.	Prédictions de l'organisation structurale-----	140
2.2.4.	Bilan des analyses <i>in silico</i> pour la séquence codante <i>IgTeCe</i> -----	142
3.	Expression hétérologue -----	145
3.1.	Préambule -----	145
3.2.	Expression dans la levure <i>Saccharomyces cerevisiae</i> -----	145
3.2.1.	Validité des systèmes d'expression contenant <i>pYE-IgTeCe</i> et <i>pYE-IgEst3</i> -----	145
3.2.2.	Caractéristiques prédites des protéines recombinantes <i>IgEst 3</i> et <i>IgTeCe</i> -----	147
3.2.3.	Expression et extraction de la protéine <i>IgEst3</i> -----	148
3.2.4.	Recherche d'activités lipolytiques pour <i>IgEst3</i> -----	149
3.2.5.	Expression et extraction de la protéine <i>IgTeCe</i> -----	149
3.2.6.	Recherche de transcrits <i>IgTeCe</i> -----	154
3.2.7.	Recherche d'activités lipolytiques pour <i>IgTeCe</i> -----	154
3.2.8.	Bilan des tests d'expression dans la levure -----	155
3.3.	Expression dans la bactérie <i>Escherichia coli</i> -----	155
3.3.1.	Clonage dans le plasmide <i>pJET1.2</i> -----	155
3.3.2.	Sous-clonage dans les vecteurs d'expression-----	158
3.3.3.	Expression hétérologue de la protéine <i>IgTeCe</i> -----	158
3.3.4.	Chromatographie d'affinité -----	162
3.4.	Recherche de transcrits pour <i>IgTeCe</i> après expression dans la bactérie -----	165
4.	Recherche d'activité pour <i>IgTeCe</i> après expression dans la bactérie -----	167
4.1.	Hydrolyse de substrats chromophores-----	167
4.2.	Hydrolyse de substrats gélosés -----	167
4.3.	Analyse de l'activité thioestérase-----	168
5.	Discussion générale -----	171
5.1.	Evolution de prédictions des fonctions d' <i>IgEst3</i> et d' <i>IgTeCe</i> -----	172
5.2.	Expression hétérologue d' <i>IgEst3</i> et d' <i>IgTeCe</i> -----	175
5.2.1.	Résultats de l'expression dans la levure <i>Saccharomyces cerevisiae</i> -----	175
5.2.2.	Résultats de l'expression d' <i>IgTeCe</i> dans la bactérie <i>Escherichia coli</i> -----	177
	<i>Conclusion et perspectives</i> -----	185
	<i>Annexes</i> -----	191
	<i>Bibliographie</i> -----	207

LISTE DES FIGURES

Figure 1: Hydrolyse enzymatique d'une liaison ester.....	9
Figure 2: Action de la lipase.....	10
Figure 3: Réactions de synthèse catalysées par les lipases (Gandhi, 1997)	10
Figure 4: Réaction d'hydrolyse d'un acyl coenzyme A par la thioestérase.....	11
Figure 5: Sites d'action des phospholipases sur les phospholipides.....	11
Figure 6: Nombre d'entrées dans la PDB pour les groupes d'enzymes lipolytiques choisis	12
Figure 7: Représentation schématique d'une hydrolase de type (Ollis <i>et al.</i> , 1992)	13
Figure 8: Représentation schématique de l'hydrolyse d'une liaison ester (Beer <i>et al.</i> , 1996)	14
Figure 9: Représentation des feuillets β et hélices α caractéristiques avec positionnement de la séquence pentapeptidique consensus (Derewenda, 1994; Kourist <i>et al.</i> , 2010).....	14
Figure 10: Conformation prise par le volet amphiphile sans interface (forme fermée) et à l'interface eau/lipide (forme ouverte) (Derewenda <i>et al.</i> , 1994; Palomo <i>et al.</i> , 2003)	15
Figure 11: Modèle cinétique d'action d'une enzyme soluble sur un substrat insoluble.	16
Figure 12: Hydrolyse de la triacétine par la carboxylestérase du foie de cheval et la lipase pancréatique de porc (Verger, 1976)	16
Figure 13: Schématisation des structures d'isozymes de phospholipase C chez la souris.	25
Figure 14: Chromatographie sur couche mince des lipides totaux d' <i>Isochrysis galbana</i> pour la mise en évidence de l'activité lipolytique endogène (Godet S, Thèse 2008).....	53
Figure 15: Voie de synthèse supposée pour les AGPIs dans les microalgues (Shi <i>et al.</i> , 2012).	54
Figure 16: Schématisation du facteur α	66
Figure 17: Schéma de base d'un vecteur d'expression d' <i>Escherichia coli</i> (European Molecular Biology Laboratory)	73
Figure 18: Schématisation d'un vecteur d'expression chez <i>Saccharomyces cerevisiae</i>	75
Figure 19: Illustration de la place des thioestérases dans la synthèse des acides gras (Zhang <i>et al.</i> , 2011)	82
Figure 20: Le vecteur d'expression <i>pYES2.1</i> (Invitrogen™)	87
Figure 21: Clonage selon la technique à la TopoIsomérase (d'après Invitrogen™)	88
Figure 22: Le plasmide de clonage <i>pJET1.2</i> (Fermentas™).....	88
Figure 23: Le vecteur d'expression <i>pET28a+</i>	89
Figure 24: Le vecteur d'expression <i>pET52b+</i>	89
Figure 25: Séquences codante et protéique d' <i>IgEst3</i>	120
Figure 26: Alignement des séquences homologues obtenues dans la banque PDB.	126
Figure 27: Composition des deux groupes d'enzymes obtenus par le logiciel PipeAlign.....	129
Figure 28: Prédictions des structures secondaires d' <i>IgEst3</i>	130
Figure 29: Organisation structurale de la triade catalytique d' <i>IgEst3</i> et de la carboxylestérase de <i>Francisella tularensis</i> . .	131
Figure 30: Modélisation tridimensionnelle complète d' <i>IgEst3</i>	133
Figure 31: Séquences codante et protéique d' <i>IgTeCe</i>	134

Figure 32: Alignement des séquences homologues obtenues dans la banque de données PDB.	138
Figure 33: Composition des deux groupes d'enzymes obtenus par le logiciel PipeAlign.....	140
Figure 34: Prédiction de la structure secondaire d' <i>IgTeCe</i>	141
Figure 35: Organisation structurale de la triade catalytique d' <i>IgTeCe</i> et de la lysophospholipase humaine [<i>HsLYPLIS</i>].	142
Figure 36: Prédiction de la structure de la protéine <i>IgTeCe</i>	143
Figure 37: Modélisation « globale » de l'ensemble de la protéine <i>IgTeCe</i>	144
Figure 38: Electrophorèse sur gel d'agarose d'ADN plasmidiques issus de la levure.	146
Figure 39: Séquence partielle du plasmide <i>pYES2.1</i>	146
Figure 40: Electrophorèse sur gel d'agarose des fragments amplifiés par PCR avec les amorces « K7EXTRACT ».	147
Figure 41: Electrophorèse sur gel d'agarose des fragments amplifiés spécifiques de chaque gène <i>IgEst3</i> et <i>IgTeCe</i>	147
Figure 42: Electrophorèse sur SDS-PAGE après chromatographie d'affinité du témoin LacZ (A) et d' <i>IgEst3</i> (B).....	148
Figure 43: Electrophorèse sur SDS-PAGE 12% des extraits de protéines.	150
Figure 44: A) Electrophorèse sur SDS-PAGE 12 % des extraits protéiques déposés pour le Western blot. B) Western blot des extraits protéiques.	150
Figure 45: Electrophorèse sur SDS-PAGE 12 % des différentes fractions de chromatographie d'affinité pour <i>IgTeCe</i> après induction de 24 heures.	152
Figure 46: Comparaison entre le témoin LacZ et <i>IgTeCe</i>	152
Figure 47: Analyse des fractions de chromatographie et des extraits protéiques obtenus pour le témoin LacZ et <i>IgTeCe</i> avec une induction de 60 heures. A) SDS-PAGE 12 % B) Western blot.	153
Figure 48: Electrophorèse sur gel d'agarose pour la recherche de transcrits <i>IgTeCe</i>	154
Figure 49: Test d'hydrolyse sur milieu gélosé avec 10 % de tributyrine.	155
Figure 50: Electrophorèse sur gel d'agarose pour les amplifications de la séquence codante <i>IgTeCe</i>	156
Figure 51: Illustration du produit de clonage d' <i>IgTeCe</i> obtenu et inséré dans pJET.....	157
Figure 52: Western blot <i>pET28-IgTeCe</i> / Rosetta™ 2.....	159
Figure 53: Western blot après extraction à 8M urée.	159
Figure 54: Validation de l'ensemble des constructions par Western blot.....	160
Figure 55: Western blot du test de différentes températures d'expression.....	162
Figure 56: SDS-PAGE et Western blot des différentes fractions obtenues par chromatographie d'affinité.	164
Figure 57: Résultat de l'étape de PCR pour la recherche de transcrits à 37°C.	165
Figure 58: Résultat de l'étape de PCR lors de la recherche de transcrits à 16°C.	166
Figure 59: Activité lipolytique sur milieu gélosé contenant une émulsion de trioléine et l'indicateur rhodamine.....	168
Figure 60: Composition du contenu en acides gras pour la culture à 16 °C.	169
Figure 61: Composition du contenu en acides gras pour la culture à 37 °C.	170
Figure 62: Vue simplifiée de la synthèse d'acides gras chez la bactérie <i>Escherichia coli</i> (Zhang <i>et al.</i> , 2011).....	179
Figure 63: Vue simplifiée des compartiments intervenant dans la synthèse des lipides chez les microalgues (Radakovits <i>et al.</i> , 2010).....	180

LISTE DES TABLEAUX

Tableau 1: Présentation des familles de thioestérases et de leurs spécificités de substrat.....	19
Tableau 2: Présentation des principales méthodes de production de lipases.....	37
Tableau 3: Données générales de la banque de données LED.....	49
Tableau 4: Quelques motifs retrouvés dans les enzymes lipolytiques à travers la base de données ProSite	51
Tableau 5: Annotations des séquences présentes dans la base de données d'EST d' <i>I. galbana</i>	55
Tableau 6: Adaptation du temps d'incubation en fonction de la température de culture (Terpe, 2006).....	59
Tableau 7: Code génétique standard.....	61
Tableau 8: Codons les plus rares chez <i>E. coli</i>	62
Tableau 9: Principales souches d'expression d' <i>E. coli</i>	64
Tableau 10: Principaux promoteurs employés avec <i>Escherichia coli</i>	70
Tableau 11: Promoteurs utilisés avec la levure <i>Saccharomyces cerevisiae</i>	71
Tableau 12: Caractéristiques générales de quelques vecteurs d'expression dans <i>Escherichia coli</i>	74
Tableau 13: Différents vecteurs d'expression, en gris le vecteur utilisé lors de cette étude.....	75
Tableau 14: Principaux critères de choix d'un organisme d'expression	76
Tableau 15: Présentation de quelques études de lipases par expression hétérologue	77
Tableau 16: Présentation de quelques études de carboxylestérases par expression hétérologue.....	79
Tableau 17: Présentation de quelques études de thioestérases par expression hétérologue	81
Tableau 18: Présentation des autres activités lipolytiques	82
Tableau 19: Antibiotiques utilisés.....	90
Tableau 20: A)Composition et B)programme pour la réaction PCR.....	93
Tableau 21: Amorces utilisées pour déterminer la présence de l'insert	94
Tableau 22: Séquences des amorces utilisées pour caractériser et cloner les gènes <i>IgTeCe</i> et <i>IgEst3</i>	96
Tableau 23: Amorces pour le séquençage dans le plasmide <i>pJET1.2</i>	99
Tableau 24: Amorces pour l'ajout de sites de restriction à <i>IgTeCe</i>	100
Tableau 25: Sites de restriction pour le sous-clonage en plasmide d'expression.....	100
Tableau 26: Amorces pour le séquençage dans les plasmides d'expression.....	101
Tableau 27: Conditions d'expression testées	103
Tableau 28: Usages spécifiques du kit d'extraction d'ARN total Promega	104
Tableau 29: Amorces PCR pour la RT-PCR	105
Tableau 30: Tampons de lyse utilisés pour l'extraction de protéines de levure	105
Tableau 31: Différents tampons de lyse utilisés sur les cellules d' <i>E. coli</i>	106
Tableau 32: Composition des gels et tampons nécessaires à l'électrophorèse de protéines	107
Tableau 33: Marqueurs de tailles utilisés pour les électrophorèses de protéines.....	107
Tableau 34: Tampons utilisés pour la chromatographie d'affinité	110
Tableau 35: Programmes d'alignement.....	122

Tableau 36: Banques de données utilisées pour l'alignement de la séquence <i>IgEst3</i>	123
Tableau 37: Liste des alignements retenus pour la protéine <i>IgEst3</i>	123
Tableau 38: BLASTP d' <i>IgEst3</i> contre les séquences des banques NR.....	124
Tableau 39: BLASTP d' <i>IgEst3</i> contre les séquences de la banque RefSeq.	124
Tableau 40: BLASTP d' <i>IgEst3</i> contre les séquences de la banque RefSeq et issues de plantes et d'algues.	125
Tableau 41: BLASTP d' <i>IgEst3</i> contre les séquences de la banque SwissProt.	125
Tableau 42: BLASTP d' <i>IgEst3</i> contre les séquences de la banque PDB.....	126
Tableau 43: DELTA-BLAST d' <i>IgEst3</i> contre les séquences des banques NR.....	127
Tableau 44: DELTA-BLAST d' <i>IgEst3</i> contre les séquences des banques NR issues de plantes.	127
Tableau 45: DELTA-BLAST d' <i>IgEst3</i> contre les séquences des banques NR issues d'algues.	128
Tableau 46: PHI-BLAST d' <i>IgEst3</i> contre les séquences des banques NR sans restriction d'organismes, l'alignement tient compte de la présence du pentapeptide GX SXG.	128
Tableau 47: Pourcentages d'identité et de similarité des séquences ayant permis l'annotation du cluster ISL 00006895 dans la base de données TBestDB par rapport à la séquence protéique complète <i>IgEst3</i>	132
Tableau 48: Liste des alignements retenus pour la protéine <i>IgTeCe</i>	135
Tableau 49: BLASTP d' <i>IgTeCe</i> contre les séquences des banques NR.	136
Tableau 50: BLASTP d' <i>IgTeCe</i> contre les séquences de la banque RefSeq.	136
Tableau 51: BLASTP d' <i>IgTeCe</i> contre les séquences de la banque RefSeq issues d'algues (macro- ou micro-).....	137
Tableau 52: BLASTP d' <i>IgTeCe</i> contre les séquences de la banque Swissprot.	137
Tableau 53: BLASTP d' <i>IgTeCe</i> contre les séquences de la banque PDB.	137
Tableau 54: DELTA-BLAST d' <i>IgTeCe</i> contre les séquences des banques NR.	138
Tableau 55: DELTA-BLAST d' <i>IgTeCe</i> contre les séquences des banques NR issues de plantes.....	138
Tableau 56: DELTA-BLAST d' <i>IgTeCe</i> contre les séquences des banques NR issues d'algues.....	139
Tableau 57: PHI-BLAST d' <i>IgTeCe</i> contre les séquences des banques NR sans restriction d'organismes, l'alignement tient compte de la présence du pentapeptide GX SXG.	139
Tableau 58: Pourcentage d'identité et de similarité des séquences ayant permis l'annotation du cluster ISL00007519 dans la base de données TBestDB par rapport à la séquence protéique complète <i>IgTeCe</i>	142
Tableau 59 : Caractéristiques des protéines natives et recombinantes.	147
Tableau 60: Comparaison des modifications apportées à la protéine <i>IgTeCe</i> sous sa forme recombinée.	158
Tableau 61: Résultats extraits par le logiciel PredAlgo pour la localisation prédictive de la protéine <i>IgTeCe</i>	181
Tableau 62: Résultats extraits par le logiciel PredAlgo pour la localisation prédictive de la protéine <i>IgEst3</i>	182
Tableau 63: Prédiction pour la localisation cellulaire d' <i>IgLip</i> et d' <i>IgEst2</i>	182

LISTE DES ABREVIATIONS

AA : Acides Aminés

ADN : Acide DésoxyriboNucléique

ADNc : Acide DésoxyriboNucléique complémentaire

ARN : Acide RiboNucléique

ARNm : Acide RiboNucléique messenger

ARNt : Acide RiboNucléique de transfert

AGPI : Acides Gras PolyInsaturés

BLAST : Basic Local Alignment Search Tool

CPG: Chromatographie en Phase Gazeuse

Da : Dalton

dNTP : désoxyriboNucléotide TriPhosphate

DHA : Acide DocosaHexaénoïque

EC : Enzyme Commission

EST: Expressed Sequence Tag

ESTHER : ESTerases and alpha/beta Hydrolase Enzyme and Relatives

IgTeCe : *Isochrysis galbana* Thioestérase Carboxylestérase

IgEst3 : *Isochrysis galbana* Ester hydrolase 3

LED : Lipase Engineering Database

MELDB : Microbial Esterase Lipase DataBase

MG : MonoGlycéride

NR : Non Redundant protein sequences

pb : paire de bases

PCR : Polymerase Chain Reaction

PDB : Protein Data Bank

Pfam: Protein Families database

RT-PCR : Reverse-Transcriptase Polymerase Chain Reaction

SDS-PAGE : Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresys

CODE DES ACIDES AMINES

Nom	Abréviation en trois lettres	Code à une lettre
Glycine	Gly	G
Alanine	Ala	A
Valine	Val	V
Leucine	Leu	L
Isoleucine	Ile	I
Sérine	Ser	S
Thréonine	Thr	T
Cystéine	Cys	C
Méthionine	Met	M
Aspartate	Asp	D
Asparagine	Asn	N
Glutamate	Glu	E
Glutamine	Gln	Q
Lysine	Lys	K
Arginine	Arg	R
Histidine	His	H
Phénylalanine	Phe	F
Tyrosine	Tyr	Y
Tryptophane	Trp	W
Proline	Pro	P

INTRODUCTION

INTRODUCTION

Les enzymes lipolytiques constituent un groupe d'enzymes qui catalysent la dégradation des lipides et de façon plus générale l'hydrolyse des esters d'acides gras. Elles appartiennent à la superfamille des hydrolases à repliement α/β et sont présentes dans l'ensemble du monde vivant. De par leurs propriétés physico-chimiques et catalytiques, elles représentent aujourd'hui l'un des enjeux majeurs des biotechnologies. En effet, de nombreux produits font appel à ces enzymes dans leur processus de fabrication. Elles interviennent pour la production d'aliments, de médicaments ou encore de papier. Elles ont également pris une place importante dans des applications domestiques. Elles sont souvent plus économiques et génèrent moins de déchets que les transformations chimiques. Ces différentes applications sont à l'origine de la caractérisation structurale et fonctionnelle d'un grand nombre d'enzymes lipolytiques et plus particulièrement de lipases. De plus, l'intérêt accru des enzymes lipolytiques oriente la recherche vers la découverte d'enzymes plus performantes et plus spécifiques que celles actuellement connues.

Les microalgues sont des organismes eucaryotes, unicellulaires et photosynthétiques présents dans les écosystèmes aquatiques voire terrestres humides. Elles sont caractérisées par des teneurs et des compositions en acides gras originales. En effet, certaines espèces présentent des teneurs élevées en acides gras poly-insaturés à longue chaîne de la famille des oméga 3 comme l'acide docosahexaénoïque (C22 :6 ω 3, DHA) connu pour son intérêt dans les domaines médical et nutritionnel. La présence de ces acides gras particuliers permet d'envisager l'existence d'enzymes lipolytiques offrant des propriétés originales. De plus, des études ont montré que la composition et les teneurs en acides gras de certaines microalgues étaient modifiées en situation de stress (température, lumière, teneur en azote ou encore salinité). En réponse à ces changements environnementaux, ces microalgues produisent des acides gras mono- et poly-insaturés dont le métabolisme peut être mis en corrélation avec l'expression de différents gènes codant des enzymes lipolytiques.

L'expression hétérologue est appliquée depuis de nombreuses années pour produire et caractériser les enzymes lipolytiques. Différents outils de biologie moléculaire permettent d'isoler la séquence codante d'intérêt et de la transférer depuis l'organisme d'origine vers l'organisme d'expression. La variété des vecteurs d'expression utilisés pour exprimer un gène recombinant permet une large application aux différentes cellules hôtes. L'association d'un vecteur à un hôte d'expression permet de

définir le système d'expression. Le choix de ce système constitue l'une des étapes critiques dans le processus de production des protéines hétérologues. Ce choix est principalement guidé par les modifications que la protéine doit subir pour être biologiquement active. Plus la protéine est complexe, plus le système d'expression doit être évolué. Il faut également tenir compte des homologies de séquences des protéines à exprimer, afin de choisir le système d'expression le plus proche des systèmes exprimant ce type de protéine. De même, il est important de connaître les compartiments cellulaires dans lesquels la protéine d'intérêt se replie ou exerce son activité biologique naturelle. Ces informations peuvent être essentielles pour définir les conditions expérimentales (compatibilité physico-chimique, choix des expérimentations).

Chez les microalgues, peu d'enzymes impliquées dans le métabolisme des lipides ont été isolées et caractérisées par expression hétérologue. Les principales études ont porté sur des élongases et des désaturases en privilégiant la levure *Saccharomyces cerevisiae* et la bactérie *Escherichia coli* comme hôtes d'expression.

Dans ce contexte, l'équipe CHIMIMAR du laboratoire MMS de Laval a choisi de s'intéresser aux enzymes lipolytiques de la microalgue *Isochrysis galbana* (Parke). Cette microalgue fait partie du règne des *Chromalveolata*, dans l'embranchement des haptophytes. Cette espèce appartient à l'ordre des *Isochrysidaceae* dans la classe des Prymnesiophycées. Elle est retrouvée essentiellement dans l'hémisphère nord (Espagne, Royaume-Uni, Etats-Unis). La particularité d'*Isochrysis galbana* est essentiellement liée à sa composition en acides gras dont sept sont présents dans des proportions comprises entre 8 % et 20 % : C14:0, C16:0, C18:1 ω 9, C18:3 ω 3, C18:4 ω 3, C20:4 ω 6 et C22:6 ω 3 (DHA). Chez cette microalgue, une activité lipolytique endogène a été soupçonnée au laboratoire par la présence d'acides gras libres dans les lipides totaux extraits de la microalgue (Poisson L, Thèse 1999). D'autres travaux avaient par la suite mis en évidence une activité endogène spécifique des phospholipides et glycolipides (Devos M, Thèse 2005). La présence d'acides gras particuliers a de plus suggéré la présence d'enzymes lipolytiques originales. Pour caractériser ces enzymes, deux choix sont possibles, soit explorer le protéome et extraire l'enzyme lipolytique d'intérêt, soit isoler les gènes putatifs et les cloner pour réaliser une expression hétérologue. Le laboratoire a choisi cette seconde stratégie. Une fois la protéine purifiée, elle peut alors être analysée et son activité enzymatique déterminée.

L'objectif de notre étude est de réaliser l'expression et la caractérisation de deux gènes codant des enzymes lipolytiques. Après avoir prédit *in silico* l'activité de chaque enzyme, ce travail de thèse se focalisera sur leur production, leur purification et leur caractérisation biochimique. Pour cela, deux systèmes, l'un procaryote et l'autre eucaryote seront testés, *Escherichia coli* et *Saccharomyces cerevisiae*.

Le premier chapitre de ce manuscrit consiste en une synthèse bibliographique. Le second chapitre décrit les techniques utilisées tout au long du travail de thèse. Le troisième chapitre, consacré aux résultats obtenus et à leurs interprétations, présentera dans un premier temps l'étude bioinformatique des deux séquences isolées de la microalgue *Isochrysis galbana*. Dans un second temps, l'expression hétérologue de ces deux séquences dans la levure et la bactérie sera développée. Enfin, l'activité enzymatique des protéines recombinantes sera recherchée. Ce chapitre se terminera par une discussion générale. Une conclusion complétée par des perspectives achèvera ce manuscrit.

CHAPITRE I : SYNTHESE BIBLIOGRAPHIQUE

CHAPITRE I : SYNTHESE BIBLIOGRAPHIQUE

1. Les enzymes lipolytiques

1.1. Définitions

Les enzymes lipolytiques sont largement répandues dans le monde vivant où elles ont un rôle physiologique important dans le métabolisme des graisses (Jaeger *et al.*, 1994). Ces enzymes se retrouvent dans les trois règnes (Gilbert, 1993), principalement sous forme de protéines extracellulaires (Vakhlu and Kour, 2006).

Les enzymes lipolytiques sont des hydrolases de classe 3 selon la classification établie par l'Union Internationale de la Biochimie et de la Biologie Moléculaire (<http://www.chem.qmul.ac.uk/iubmb/enzyme/>). Treize sous-classes d'enzymes sont répertoriées dans cette classe et inventoriées en fonction du type de liaison clivée. La sous-classe EC 3.1 définit les enzymes qui hydrolysent une liaison ester (Figure 1). Elles sont appelées ester hydrolases ou estérases (Villeneuve and Foglia, 1997; Jaeger and Reetz, 1998). Cette sous-classe contient différents groupes dont les hydrolases spécifiques des esters carboxyliques (EC 3.1.1) appelées aussi carboxylester hydrolases (Jaenicke, 1965). Les dénominations des différentes catégories d'enzymes lipolytiques ont été définies selon le type de lipide qu'elles hydrolysent. Il est ainsi retrouvé les thioestérases, les phospholipases et les cutinases. Les lipases (EC 3.1.1.3) et carboxylestérases (EC 3.1.1.1) sont intégrées dans la classe des carboxylester hydrolases.

Actuellement, la définition d'une enzyme lipolytique selon l'« Enzyme Commission » regroupe de nombreux groupes d'enzymes et rend très complexe la classification d'une nouvelle enzyme. Ali *et al.* (2012) ont montré que la définition des enzymes lipolytiques doit évoluer sur de nouveaux critères d'identification permettant de différencier les estérases ayant ou non pour substrat des lipides. Cependant, la littérature utilise le terme « estérase » classiquement pour désigner les carboxylestérases par opposition aux lipases. La suite de ce document utilise le terme de carboxylestérase pour désigner les enzymes de la classe EC 3.1.1.1.

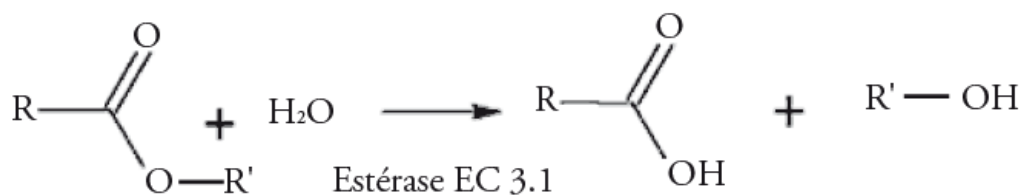


Figure 1: Hydrolyse enzymatique d'une liaison ester

Lipases et carboxylestérases

Les lipases sont actives sur des esters d'acylglycérol comme les triglycérides (Figure 2).

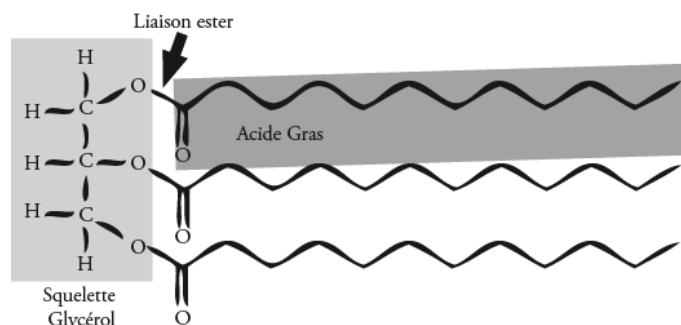


Figure 2: Action de la lipase

Les lipases catalysent préférentiellement l'hydrolyse d'esters insolubles à longue chaîne carbonée supérieure à 10 carbones comme par exemple l'acide oléique C18:1. Cette spécificité pour les esters à longue chaîne permet de différencier les lipases des carboxylestérases qui, elles, agissent sur les esters solubles à courte chaîne carbonée inférieure à 10 carbones. Les lipases sont aussi capables d'hydrolyser les substrats solubles des carboxylestérases même si ces réactions sont généralement beaucoup plus lentes. En revanche, les carboxylestérases ne peuvent pas hydrolyser des esters à longue chaîne. Les lipases sont des enzymes particulières, capables d'agir en milieu biphasique, à l'interface lipide/eau. L'activité lipolytique dépend alors de la qualité de cette interface (Hasan *et al.*, 2009). Les lipases interviennent dans deux types de réaction : l'hydrolyse et la synthèse. La figure 3 présente l'ensemble des réactions de synthèse catalysées par les lipases.

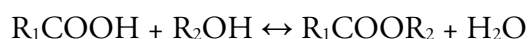
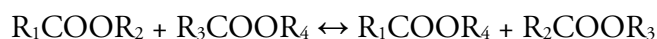
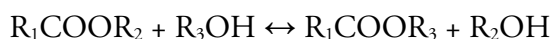
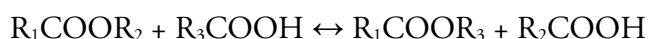
Estérification***Transestérification :******Interestérification :******Alcoolyse :******Acidolyse :***

Figure 3: Réactions de synthèse catalysées par les lipases (Gandhi, 1997)

Thioestérases

La classe identifiée EC 3.1.2, classe des thioestérases, définit les enzymes clivant la liaison ester entre un groupement carbonyle et un atome de soufre (Figure 4) et libérant un acide gras. Parmi les 27 sous-classes qui la composent, 15 ont pour substrat un dérivé d'acyl coenzyme A (CoA) (Cantu *et al.*, 2010).

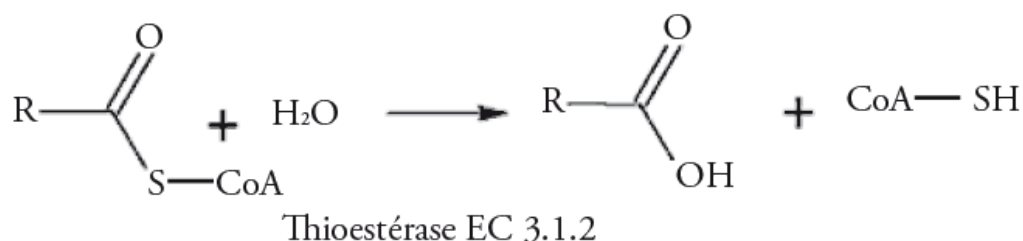


Figure 4: Réaction d'hydrolyse d'un acyl coenzyme A par la thioestérase

Phospholipases

Les phospholipases désignent plusieurs sous-groupes d'enzymes, certains libèrent des acides gras depuis un phospholipide (Figure 5) (Aloulou *et al.*, 2012). Ces enzymes sont classées selon leur site d'action. La position *sn-1* est clivée par les phospholipases A1 (PLA1) (EC 3.1.1.32), tandis que la position *sn-2* est clivée par les phospholipases A2 (PLA2) (EC 3.1.1.4), ces deux sous-groupes libèrent des acides gras. Les phospholipases C (EC 3.1.4.3, position *sn-3*) et phospholipases D (EC 3.1.4.4) sont impliquées dans le métabolisme des phospholipides mais ne libèrent pas d'acide gras.

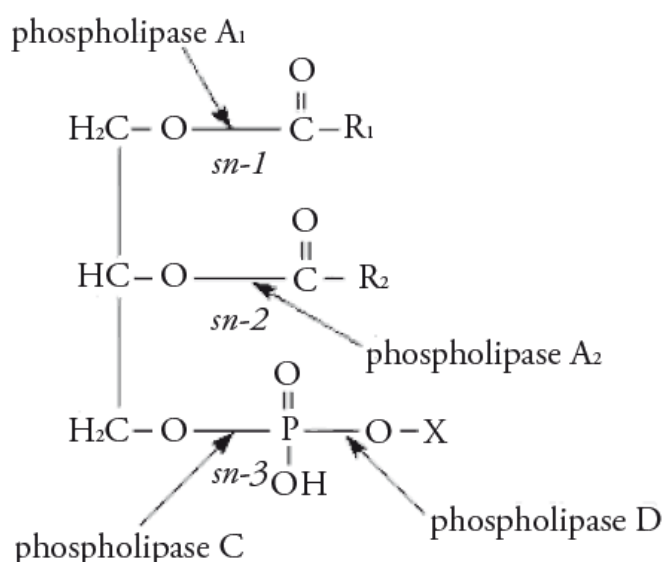


Figure 5: Sites d'action des phospholipases sur les phospholipides

Cutinases

Le groupe des cutinases (EC 3.1.1.74) est composé d'enzymes qui dégradent la cutine, un polymère d'acides gras des plantes supérieures recouvrant la face externe des feuilles et offrant une protection hydrophobe. Les cutinases ont la particularité d'être spécifiques d'acides gras à longue chaîne (C16 ou C18) *in vivo* mais de se révéler plus efficace que les lipases pour des activités d'hydrolyse d'esters solubles. Ces enzymes se définissent comme l'intermédiaire idéal entre lipases et carboxylestérases (Sorensen *et al.*, 2007).

1.2. Structure et mécanisme

La Protein Data Bank (PDB) est une base de données de structures de macromolécules biologiques. Elle est composée de structures déterminées expérimentalement soit par cristallisation ou par RMN. La recherche au sein de cette base peut être réalisée à partir d'une séquence, d'une structure ou d'une fonction recherchée. Environ 85 000 structures y sont classées parmi lesquelles 4 800 correspondent à la classification EC 3.1. Le nombre de structures a nettement augmenté depuis une quinzaine d'années notamment grâce au développement de la génomique structurale et à l'amélioration des conditions d'analyse (robotisation et automatisation). Les informations structurales disponibles sur les enzymes lipolytiques ont également augmenté en particulier pour les phospholipases et les thioestérases (<http://www ww pdb.org/stats.html>). Cette évolution, illustrée par la figure 6, permet également de rendre compte de l'intérêt croissant pour les enzymes lipolytiques au cours des douze dernières années.

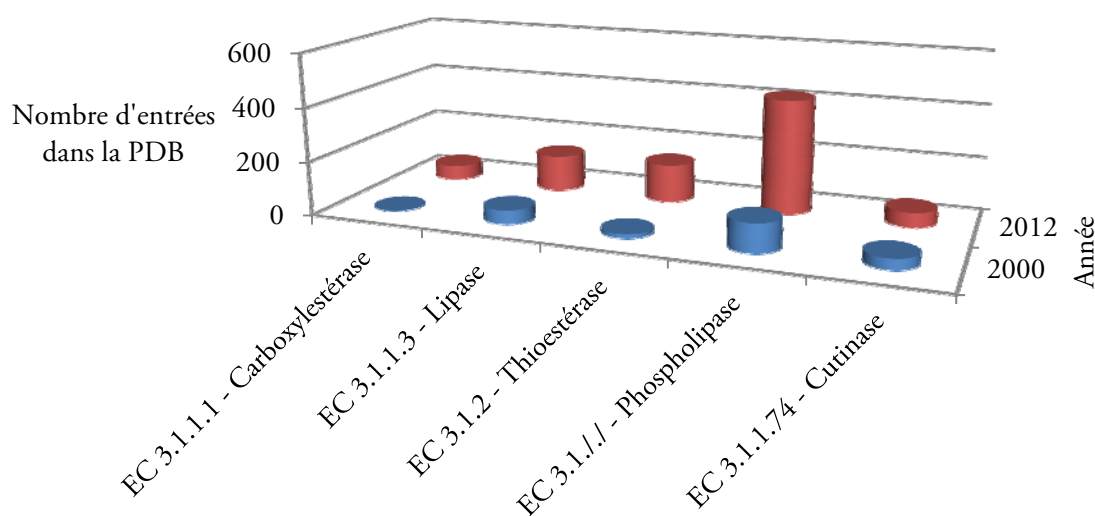


Figure 6: Nombre d'entrées dans la PDB pour les groupes d'enzymes lipolytiques choisis

1.2.1. Le système α/β hydrolase

Les hydrolases ont en commun un enchaînement de huit feuillets β connectés par six hélices α . Ce modèle est retrouvé non seulement au sein des lipases et des carboxylestérases mais également au sein d'autres groupes d'hydrolases (Figure 7). Cette organisation structurale module le positionnement du site catalytique. Dans le cas des lipases, le site catalytique est constitué autour de trois partenaires formant la triade catalytique : une sérine, une histidine et enfin un acide aminé portant au niveau de sa chaîne latérale une fonction carboxylique (Asp ou Glu) (Schmid and Verger, 1998).

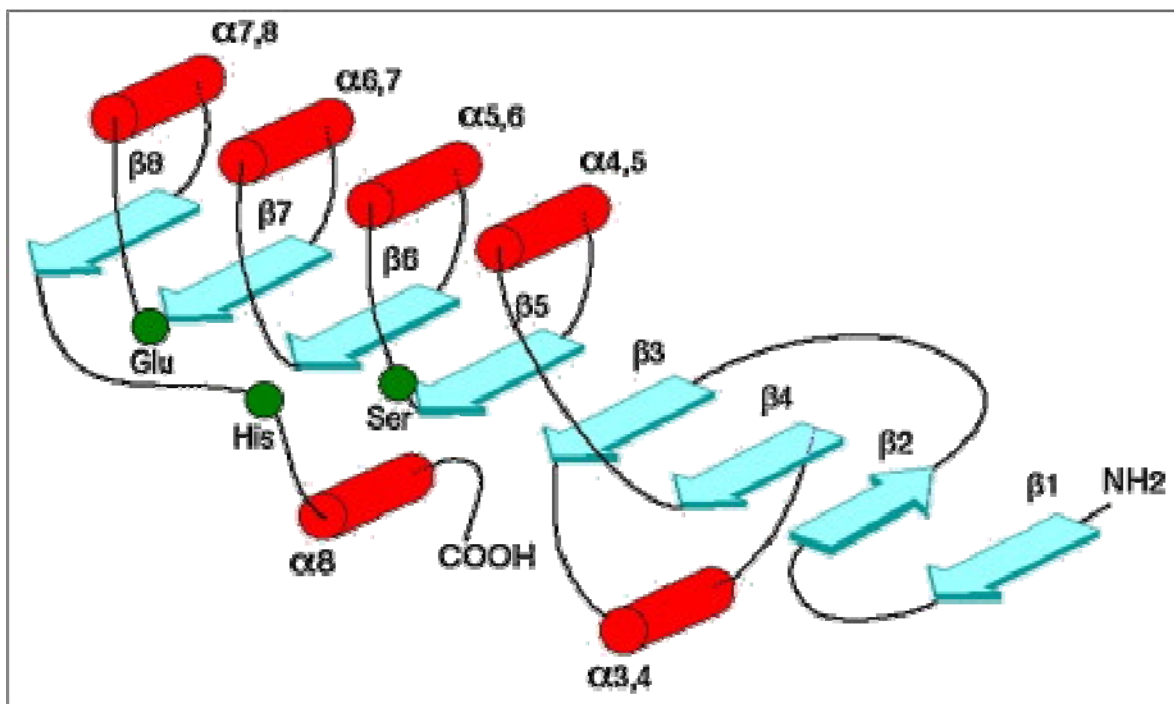


Figure 7: Représentation schématique d'une hydrolase de type (Ollis *et al.*, 1992)

Certaines lipases bien caractérisées ont permis de distinguer d'autres résidus importants pour l'activité lipolytique. Ces résidus forment une conformation appelée « trou de l'oxyanion » permettant de stabiliser l'environnement du site actif. Le mécanisme catalytique de l'hydrolyse peut être décrit en six étapes (Figure 8). Le carbone de la fonction carboxylique du substrat subit une attaque nucléophile du groupement hydroxyle de la sérine. A ce niveau, une liaison hydrogène s'est formée avec le noyau Imidazole de l'histidine qui devient alors chargé positivement. Cette charge positive est stabilisée par une charge négative d'un résidu acide (Asp ou Glu) (1) permettant la formation d'un premier oxyanion (2). Une molécule d'alcool est libérée et une acyl-enzyme est formée (3). Par la suite une molécule d'eau réalise une attaque nucléophile de l'acyl-enzyme (4). Cette seconde attaque nucléophile aboutit à la formation d'un nouvel oxyanion (5). Enfin, l'acide gras est libéré et l'enzyme retrouve sa conformation initiale (6) (Jaeger *et al.*, 1994; Misset *et al.*, 1994; Beer *et al.*, 1996).

Dans le cas de la lipase pancréatique, il a été établi que la présence de la sérine est indispensable pour obtenir une activité lipolytique. La substitution de la sérine par d'autres acides aminés dans la triade catalytique empêche en effet toute activité lipolytique. Ces mutations n'interfèrent pas dans la liaison au substrat, la sérine n'est pas impliquée dans l'interaction du site catalytique (Lowe, 1992).

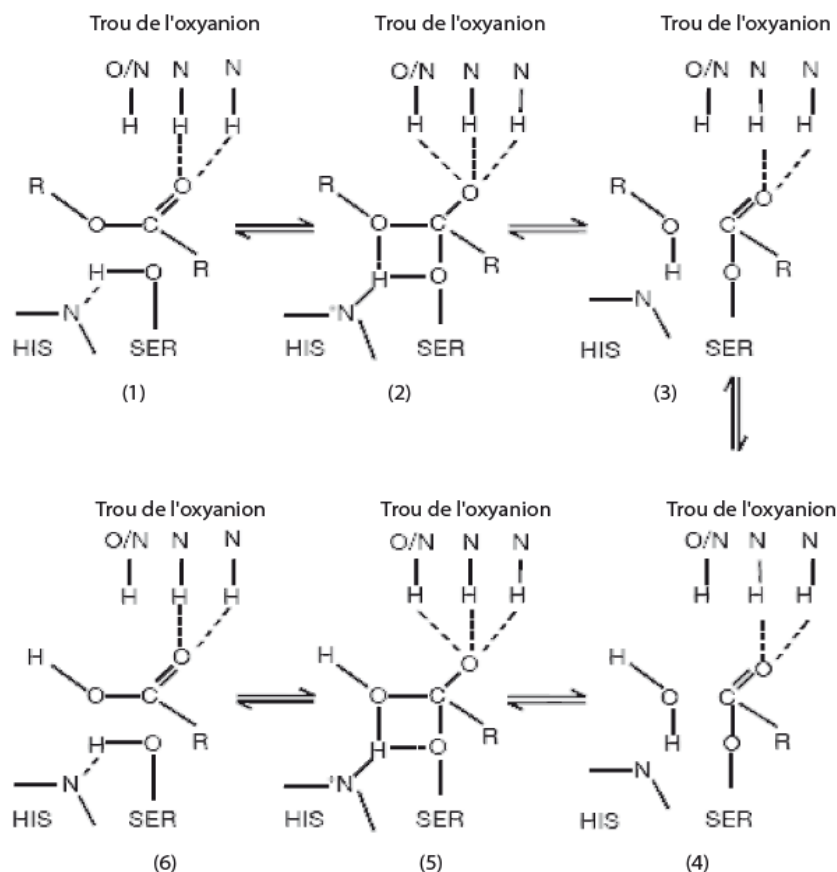


Figure 8: Représentation schématique de l'hydrolyse d'une liaison ester (Beer *et al.*, 1996)

Une autre particularité de ce site catalytique est le positionnement de la sérine dans un motif peptidique pentamérique appelé séquence consensus. Cette séquence est composée de l'enchaînement GX₁SX₂G où les X peuvent être des acides aminés d'identité différente, particularité due à la phylogénie de ce motif présent dans les protéases à sérine et les lipases (Figure 9)(Derewenda, 1994).

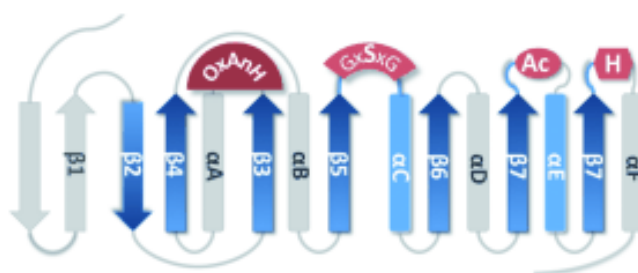


Figure 9: Représentation des feuillets β et hélices α caractéristiques avec positionnement de la séquence pentapeptidique consensus (Derewenda, 1994; Kourist *et al.*, 2010).

Contrairement aux autres hydrolases, le site actif des lipases est généralement recouvert d'une boucle peptidique formée par une hélice α amphiphile d'une quinzaine d'acides aminés qui agit comme un volet (lid) (Brzozowski *et al.*, 1991; Derewenda *et al.*, 1992) (Figure 10). L'originalité des lipases prend tout son sens par la description de cette particularité structurale car ce volet s'ouvrira uniquement à l'interface eau/lipide. Une fois ouvert, ce volet rend le site actif accessible au substrat. Dans le cas contraire, le substrat ne peut pas être en interaction avec la triade catalytique. L'activation interfaciale demeure un phénomène complexe. Thirstrup *et al.* (1994) ont ainsi montré que la lipase du ragondin malgré la présence d'un volet amphiphile ne présente pas d'activation interfaciale.

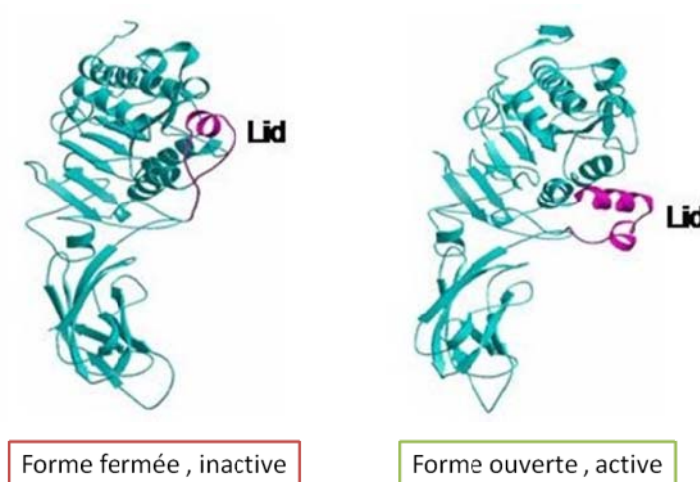


Figure 10: Conformation prise par le volet amphiphile sans interface (forme fermée) et à l'interface eau/lipide (forme ouverte) (Derewenda *et al.*, 1994; Palomo *et al.*, 2003)

1.2.1. Cinétique enzymatique des enzymes lipolytiques

Un certain nombre d'enzymes sont actives aux interfaces dans les cellules, mais ces enzymes ont des substrats « classiques », c'est-à-dire solubles et qui répondent aux principes généraux de l'enzymologie. Le substrat des lipases étant insoluble, les principes de Michaëlis et Menten ne peuvent pas être appliqués (Aloulou *et al.*, 2006). Les lipases sont activées à l'interface eau/lipide, ce phénomène est appelé activation interfaciale (Sarda and Desnuelle, 1958). Dans le cas des lipases et des phospholipases, au moins deux modèles caractérisent leur comportement aux interfaces. La première étape commune à ces deux modèles est l'adsorption de l'enzyme à l'interface eau/lipide (Benzonana and Desnuelle, 1965). La vitesse de la réaction est alors fonction de la quantité de substrat à l'aire interfaciale. Il est alors possible de définir des caractéristiques interprétables en termes de performance et de spécificité de substrat des enzymes concernées. Le modèle en figure 11 A est le modèle le plus proche de celui de Michaëlis et Menten car il n'intègre que le critère d'activation de l'enzyme. Au niveau de l'interface eau/lipide, l'enzyme est activée (E^*) et se lie à son substrat (E^*S). Un produit (P^*) est formé qui diffuse (P) dans la phase aqueuse.

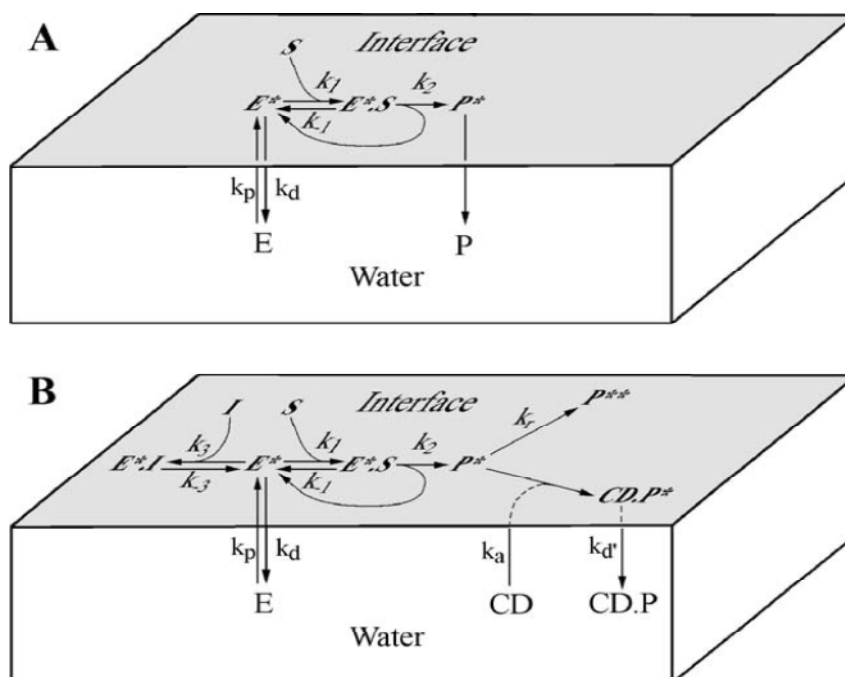


Figure 11: Modèle cinétique d'action d'une enzyme soluble sur un substrat insoluble.
Modèle Verger-De Haas (A), Modèle Verger-Panaïotov (B)

Le modèle en figure 11 B tient compte des modulations d'activité possibles à l'interface. D'une part, l'activation interfaciale rentre en jeu (E^*I). D'autre part, l'activité peut dépendre de l'organisation des produits insolubles formés (P^{**} & P^*). Enfin, la prise en charge des produits insolubles par un accepteur (CD) est également modélisée ($CD + P^* \rightarrow CD \cdot P$).

La cinétique des lipases n'est pas évidente à évaluer de par le caractère hydrophile de l'enzyme et le caractère lipophile de son substrat, l'activité prenant forme lorsque la lipase se trouve à l'interface phase aqueuse/phase lipidique (Reis *et al.*, 2009). C'est sur ce point que la différence de cinétique entre lipase et carboxylestérase se caractérise. La concentration en substrat est un paramètre limitant pour les lipases et non pour les carboxylestérases comme le montre la figure 12.

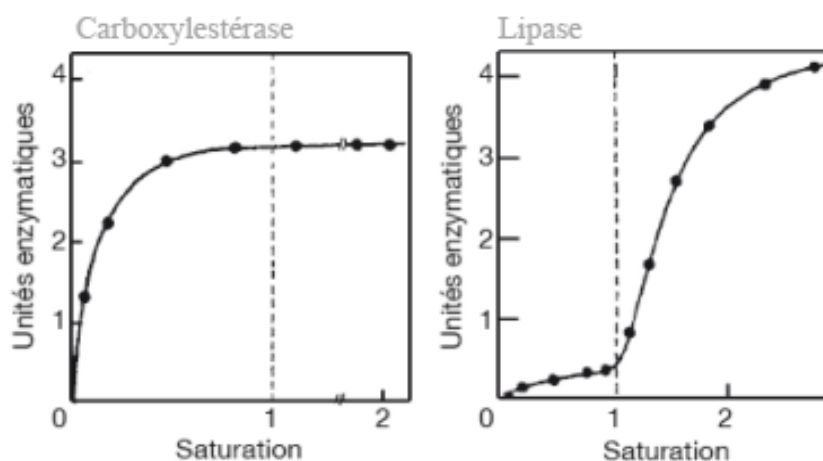


Figure 12: Hydrolyse de la triacétine par la carboxylestérase du foie de cheval et la lipase pancréatique de porc (Verger, 1976)

1.2.2. Spécificités de substrat des enzymes lipolytiques

Comme vu dans le paragraphe 1.1, la classification des hydrolases à sérine est principalement basée sur la nature du substrat. Parmi les enzymes lipolytiques, les lipases ont depuis longtemps été étudiées et caractérisées pour déterminer leur spécificité. Cette spécificité pour les acylglycérols est définie selon trois critères : régiosélectivité, typosélectivité et énantiosélectivité (Jensen *et al.*, 1983). Plus récemment, l'étude d'enzymes libérant des acides gras appartenant à la classe des thioestérases a également permis de définir des spécificités selon la nature du substrat. Ce paragraphe décrit les spécificités pour ces deux classes d'enzymes.

Spécificités des lipases

La régiosélectivité : elle définit la préférence d'une lipase pour une position donnée sur le squelette carboné du glycérol. L'enzyme a l'aptitude de distinguer les positions externes (*sn-1* et *sn-3*) de la position interne (*sn-2*). Il faut établir le rapport entre les différents types de diglycérides produits *in vitro*, afin de définir la régiosélectivité d'une lipase. Par exemple, l'hydrolyse de la trioléine peut donner trois types de produits (Anderson *et al.*, 1998) :

- Obtention d'un rapport 2/1 entre les 1,2 (2,3) diglycérides et 1,3 diglycérides : lipase non régiosélective.
- Obtention de 1,2 (2,3) diglycérides et 2 monoglycérides : lipase régiosélective pour *sn-1* et *sn-3*.
- Obtention de 1,3 diglycérides : lipase régiosélective pour *sn-2*.

La majeure partie des lipases sont régiosélectives pour les positions externes, c'est le cas pour les lipases de *Rhizopus arrizhus*, *Yarrowia lipolytica*, *Rizopus delemar* ou *Aspergillus niger*. Les lipases spécifiques de la position *sn-2* sont nettement plus marginales, elles ont été retrouvées chez *Staphylococcus aureus*, *S. simulans*, *S. xylosus* ou chez *Candida antarctica* (Pleiss *et al.*, 2000; Horchani *et al.*, 2010).

La typosélectivité : c'est l'hydrolyse spécifique d'une lipase par rapport à un type d'acide gras donné ou une famille d'acide gras donnée. La typosélectivité peut être déterminée en comparant l'activité enzymatique obtenue pour différents acides gras. Cette sélectivité se rapporte à la longueur de chaîne ou au degré d'insaturation de la chaîne carbonée à hydrolyser. Ceci est illustré, par exemple, par les isoformes de la lipase de *Candida rugosa* dont les modifications du site de liaison permettent l'adaptation à différentes longueurs de chaîne carbonée (Lotti *et al.*, 1994). L'origine des lipases peut également donner des particularités liées à l'espèce. Ainsi, les lipases animales agiront de préférence

sur les triacylglycérols alors que les lipases de certains champignons auront une affinité plus importante pour les formes avec une ou deux longues chaînes carbonées (Villeneuve and Foglia, 1997).

L'énantiosélectivité ou stéréosélectivité : il s'agit de l'hydrolyse privilégiée d'une des deux positions *sn-1* ou *sn-3* d'un triglycéride (facilement mise en évidence si elles sont occupées par des acides gras différents). Il s'agit également de l'hydrolyse sélective d'un énantiomère par rapport à un autre (cas d'un mélange racémique 1,2 diglycéride / 2,3 diglycéride). Cette spécificité permet l'utilisation de lipases lors de la synthèse de molécules chirales (Irimescu *et al.*, 2010). L'aptitude des lipases à distinguer des énantiomères a été décrite dans la synthèse de différents esters (Engel, 1992; Lutz *et al.*, 1992).

Spécificités des thioestérases

La classification de ces enzymes est basée à la fois sur la fonction enzymatique, mais également sur la nature du substrat. Les thioestérases sont classées dans la base de données ThYme (Thioester-active enzYme database). Le tableau 1 présente l'organisation des thioestérases dans cette base de données en fonction de leur structure primaire (séquence) et tertiaire (repliement). Cette classification permet également de connaître les différents substrats de cette famille d'enzyme. Elles sont classées de TE 1 à TE 23. Les enzymes des familles TE1 à TE13 hydrolysent différents types d'acyl-CoA.

La plupart des membres de ces familles disposent d'une conformation « HotDog ». Cette conformation comprend une hélice α recouverte de plusieurs feuillets β contrairement aux hydrolases α/β qui sont une succession de feuillets et d'hélices comme vu précédemment (Dillon and Bateman, 2004). Parmi les treize familles de thioestérases d'acyl-CoA, les familles TE1 à TE3 n'adoptent pas la conformation « HotDog » et seule la famille TE2 prend la conformation d' α/β hydrolase.

Les enzymes TE2 sont sélectives des palmitoyl-CoA notamment. La longueur de chaîne est un critère de sélection de substrat pour les enzymes TE4 : moyenne et longue chaîne d'acide gras et TE5 : longue chaîne. Les enzymes composant les familles TE14 à TE19 clivent la liaison entre l'acyl et une protéine de transport d'acyl (ACP), la longueur de la chaîne carbonée est également prépondérante pour la caractérisation. Les familles TE 20 et TE 21 clivent la liaison entre l'acyl et d'autres types de protéines. Par exemple pour la famille TE 20, la coupure se fait entre une chaîne de type palmitique et le résidu cystéine de la protéine. Les enzymes de la famille TE21 permettent également cette hydrolyse et peuvent avoir une activité carboxylestérase (Duncan and Gilman, 1998). Enfin, les

enzymes des groupes TE 22 et TE 23 sont plus spécifiques des liaisons entre acyl et glutathion et ses dérivés.

Famille	Nom de gène ou de famille d'enzyme	Spécificités	Fonction	Organisme producteur	Repliement
TE1	Ach1	Acetyl-CoA	Acyl-CoA hydrolase	ABE	NagB
TE2	Acot1–Acot6, BAAT thioesterase	Palmitoyl-CoA, bile-acid-CoA	Acyl-CoA hydrolase	ABE	α/β -Hydrolase
TE3	tesA, acyl-CoA thioesterase I, protease I, lysophospholipase L1	Acyl-CoA à chaîne moyenne à longue	Acyl-CoA hydrolase	ABE	Flavodoxin-like
TE4	tesB, acyl-CoA thioesterase II, Acot8	Acyl-CoA à chaîne moyenne à longue, palmitoyl-CoA, choloyl-CoA	Acyl-CoA hydrolase	BE	HotDog
TE5	tesC (ybaW), acyl-CoA thioesterase III	Acyl-CoA à chaîne longue, 3,5-tetradecadienoyl-CoA	Acyl-CoA hydrolase	BE	HotDog
TE6	Acot7, Acot11 (BFIT, Them1), Acot12 (CACH), YciA	Acyl-CoA à chaîne courte à longue, C4 - C18	Acyl-CoA hydrolase	ABE	HotDog
TE7	Acot9, Acot10	Acyl-CoA à chaîne courte à longue	Acyl-CoA hydrolase	BE	HotDog
TE8	Acot13 (Them2)	Acyl-CoA à chaîne courte à longue	Acyl-CoA hydrolase	ABE	HotDog
TE9	YbgC	Acyl-CoA à chaîne courte à longue, 4-hydroxybenzoyl-CoA	Acyl-CoA hydrolase	B	HotDog
TE10	4HBT-I	4-Hydroxybenzoyl-CoA	Acyl-CoA hydrolase	B	HotDog
TE11	4HBT-II, EntH (YbdB)	4-Hydroxybenzoyl-CoA	Acyl-CoA hydrolase	B	HotDog
TE12	DNHA-CoA hydrolase	1,4-Dihydroxy-2-naphtoyl-CoA	Acyl-CoA hydrolase	BE	HotDog
TE13	paaI, paaD	Acyl-CoA à chaîne courte à moyenne, quelques substrats hydroxyphenylacetyl-CoA	Acyl-CoA hydrolase	AB	HotDog
TE14	FatA, FatB	Acyl-ACP à chaîne courte à longue, C8 - C18	Acyl-ACP hydrolase	BE	HotDog
TE15	Thioesterase CalE7	/	Acyl-ACP hydrolase	B	HotDog
TE16	TE domain of FAS (Thioesterase I), TE domain of PKS or NRP (type I thioesterase (TE I))	Acyl-ACP à chaîne longue, divers polyketides et peptides non-ribosomaux	Acyl-ACP hydrolase	ABE	α/β -Hydrolase
TE17	TE domain of PKS	Quelques polyketides	Acyl-ACP hydrolase	B	α/β -Hydrolase
TE18	Thioesterase II, type II thioesterase (TE II)	Acyl-ACP à chaîne moyenne, divers polyketides et peptides non-ribosomaux	Acyl-ACP hydrolase	BE	α/β -Hydrolase
TE19	luxD	Myristoyl-ACP	Acyl-ACP hydrolase	B	α/β -Hydrolase
TE20	ppt1, ppt2, palmitoyl-protein thioesterase	Palmitoyl-protein	Protein-palmitoyl hydrolase	E	α/β -Hydrolase
TE21	apt1, apt2, acyl-protein thioesterase, phospholipase, carboxylesterase	//	Protein-acyl hydrolase	ABE	α/β -Hydrolase
TE22	S-formylglutathione hydrolase, esterase A, esterase D	S-Formylglutathione	Glutathione hydrolase	ABE	α/β -Hydrolase
TE23	Hydroxyglutathione hydrolase, glyoxalase II	D-Lactoylglutathione	Glutathione hydrolase	ABE	Lactamase

Tableau 1: Présentation des familles de thioestérases et de leurs spécificités de substrat.
Archées (A), Bactérie (B), Eucaryote (E) (Cantu *et al.*, 2010)

1.3. Les enzymes lipolytiques à travers le monde vivant

Les enzymes lipolytiques sont très largement répandues dans le monde vivant. Le rôle de ces enzymes est très varié d'un règne à l'autre, d'une espèce à l'autre voire d'un tissu à l'autre (Pahoja and Sethar, 2002). A ce jour, ce sont les enzymes lipolytiques de micro-organismes qui ont été les plus étudiées de part leur important intérêt industriel. D'autres sources d'enzymes lipolytiques existent dans le règne animal et végétal. Elles peuvent présenter des applications intéressantes et être exploitées (Casas-Godoy *et al.*, 2012). Les enzymes lipolytiques ont des caractéristiques variables selon les règnes. Par exemple, la zone de pH optimum de l'activité, varie entre 4 et 9 pour les enzymes lipolytiques de plantes alors qu'elle est comprise entre 5,5 et 8,5 chez les animaux et entre 6 et 10 chez les micro-organismes. Ce sous-chapitre expose les principales enzymes lipolytiques dans les règnes végétal, animal et microbien.

1.3.1. Les enzymes lipolytiques végétales

Les principales enzymes lipolytiques de plantes se classent en trois catégories en fonction de leur origine. Elles peuvent être issues de plantes oléagineuses, de plantes céréalières et de plantes laticifères (produisant du latex). Ces enzymes se regroupent alors en quatre groupes principaux : les lipases, les phospholipases A, les phospholipases C et D, et les carboxylestérases.

Certaines enzymes ne sont pas spécifiques vis-à-vis d'un tissu et interviennent de manière générale dans le métabolisme. La plus grande concentration d'enzymes lipolytiques demeure au niveau des graines où les triglycérides sont stockés. Les enzymes lipolytiques présentes dans ces structures vont hydrolyser les triglycérides et libérer des acides gras. Ces acides gras seront la source d'énergie principale lors des étapes de germination et des premiers stades de croissance de la plante.

Les premières enzymes lipolytiques de plantes ont été étudiées au début du XX^e siècle, mais il faut attendre les années soixante pour que des études plus poussées apparaissent. Elles sont réalisées sur la lipase acide du ricin, *Ricinus communis*, en 1962. La plupart des lipases connues dans les graines sont des lipases alcalines, tandis que celle du ricin a la particularité d'agir à un pH optimum de 4,5. La seconde propriété importante de cette enzyme est le déclenchement particulier de son activité. En effet, celle-ci est détectée lors des deux premiers jours de croissance alors que la sénescence n'a pas encore commencé. L'activité de la plupart des autres lipases retrouvées dans les graines est observée au moins au troisième jour de croissance. Chez le ricin, la lipase acide semble à son maximum d'activité avant l'hydrolyse des triglycérides. Une autre lipase dont le pH optimum est neutre est présente chez cette espèce. Son activité est similaire aux lipases habituellement rencontrées dans les

graines des autres plantes. Cette lipase semble intervenir pour l'hydrolyse des triglycérides au moment de la croissance de la plante (Ory and Altschul, 1962; Ory *et al.*, 1969; Muto and Beevers, 1974). Le ricin possède également une protéine végétale ayant une activité lipolytique et le motif GX SXG sans être une lipase, la ricine. C'est une glycoprotéine parmi les plus toxiques au monde. Elle intervient notamment dans les mécanismes de défense des plantes dans leur environnement (Helmy *et al.*, 1999; Lombard *et al.*, 2001).

Dans les années 1995-2000, l'abondance des séquençages de transcrits et le séquençage du génome d'*Arabidopsis thaliana* ont mis en évidence un grand nombre de lipases putatives. Ces annotations ont été uniquement basées sur le motif GX SXG. Or de nombreuses protéines présentant ce motif n'ont pas nécessairement une activité lipolytique (cas des ester hydrolases non lipolytiques). Le motif consensus GX SXG des lipases peut en effet être présent chez d'autres enzymes. Il apparaît, par exemple, dans la chlorophyllase de la plante *Chenopodium album*, enzyme responsable de la dégradation de la chlorophylle (Tsuchiya *et al.*, 1999). De plus, ce critère est insuffisant pour d'autres enzymes n'incluant pas ce motif mais possédant une activité lipolytique. L'analyse *in silico* nécessite donc en complément une recherche biochimique de l'activité lipolytique (Li-Beisson *et al.*, 2010).

Pour caractériser les enzymes lipolytiques de plantes il est nécessaire de bien connaître les compartiments de stockage des lipides. Les triglycérides des feuilles sont stockés dans les plastoglobules. Les plastoglobules sont des corps lipidiques présents au niveau des chloroplastes, incorporant des lipoprotéines et permettant la mobilisation des triglycérides (Brehelin *et al.*, 2007; Lin and Oliver, 2008). Padham *et al.* (2007) ont montré chez *Arabidopsis thaliana* la présence à ce niveau d'une lipase. D'autres études se focalisent sur les lipases des graines des oléagineux et particulièrement sur les lipases impliquées dans l'hydrolyse des triglycérides dans les tissus de réserve (oléosomes) du colza et du maïs (Yatsu and Jacks, 1972). La caractérisation de lipases de plantes a montré l'existence de plusieurs types de lipases. Les variations concernent le pH optimum et la localisation subcellulaire. Ces lipases sont présentes au niveau des corps lipidiques, des glyoxysomes ou encore des microsomes. Chez le colza, *Brassica napus*, c'est une lipase insoluble qui a été isolée, dans une fraction moins dense que les microsomes (Weselake *et al.*, 1989).

Les phospholipases de plantes sont regroupées selon leurs rôles physiologiques et cellulaires tels que la régulation cellulaire, le changement de composition des lipides membranaires et la dégradation des lipides. Les phospholipases D sont particulièrement impliquées dans le catabolisme des

phospholipides au moment de la croissance de la plante et en réponse à un stress de l'environnement. Ces phospholipases sont variées d'un point de vue moléculaire et leur présence peut être modulée notamment par du calcium (Wang, 2000). Les phospholipases C (PLC) sont séparées en deux grandes catégories, les PLC spécifiques des phosphoinositides et les PLC sans spécificité. Les premières sont majoritairement impliquées dans la production d'inositol triphosphate et leur activité est le plus souvent une réponse à un stimulus environnemental. Les secondes participent à des mécanismes plus classiques de dégradation de phospholipides (Wang, 2001; Wang *et al.*, 2012). Les phospholipases A catalysent l'hydrolyse des phospholipides de la membrane cellulaire. Chez *Arabidopsis thaliana* elles se retrouvent en trois catégories, les protéines « patatin like », les enzymes « DAD-like » et les phospholipases A2 sécrétées. Les protéines du type « patatin like » disposent d'une diade catalytique Ser-Asp à la place de la triade catalytique classique permettant une hydrolyse des positions *sn-1* et *sn-2* des phospholipides et galactolipides. Les enzymes de déacylation « DAD-like » présentent notamment des activités de lipases (Canonne *et al.*, 2011).

De nombreuses carboxylestérases ont été identifiées. Elles représentent une grande famille d'enzymes impliquées dans de nombreux processus, notamment au niveau du métabolisme secondaire chez les végétaux supérieurs. Cette famille identifiée CXE a été principalement décrite au sein d'*Arabidopsis thaliana* et d'*Oryza sativa* (Yu *et al.*, 2002). Cette famille conserve un motif peptidique et une structure secondaire correspondant aux enzymes de la famille des α/β hydrolases. La caractérisation d'une majorité des carboxylestérases de plantes n'est pas aussi avancée que leurs homologues d'origine animale. Néanmoins, un certain nombre d'informations a été déterminé, notamment la localisation dans les graines, fruits et feuilles. L'isolement et la caractérisation des gènes de carboxylestérases chez la plante modèle *Arabidopsis thaliana* demeurent récents. Ils ont permis d'identifier un certain nombre de rôles physiologiques des carboxylestérases au sein de cette plante. Aucune enzyme n'a été clairement identifiée pour l'hydrolyse de lipides mais elles interviennent dans l'initiation de signalisations cellulaires. Les carboxylestérases apparaissent également dans des processus de dégradation de certains composés comme les xénobiotiques et les esters (Marshall *et al.*, 2003). Plus récemment, le gène d'une carboxylestérase d'*Actinidia eriantha* a été isolé et cloné dans une bactérie. La protéine obtenue a été utilisée pour obtenir la première structure de carboxylestérase de plante. L'analyse de cette structure a conforté l'identification de l'enzyme dans la famille des α/β hydrolases et a également permis l'identification des acides aminés impliqués dans le site catalytique.

En complément de l'étude structurale, une caractérisation fonctionnelle a été réalisée confirmant la spécificité pour les faibles longueurs de chaînes (Ileperuma *et al.*, 2007).

Les thioestérases de plantes jouent un rôle important dans la composition des triglycérides des tissus de stockage. De ce fait, leurs caractéristiques sont largement étudiées. Elles sont considérées comme une famille d'enzymes à part dans les familles de thioestérases. En effet, peu de similarités de séquences existent avec les thioestérases des procaryotes et des animaux. La seconde différence majeure est le manque de sensibilité aux inhibiteurs connus des thioestérases bactériennes et animales. Ces inhibiteurs ciblent les protéases à sérine. L'analyse d'une séquence de thioestérase d'*Arabidopsis thaliana* montre que les trois acides aminés impliqués dans le site catalytique sont la cystéine 264, l'histidine 229 et l'asparagine 227. Une cystéine remplace donc la sérine dans la triade catalytique. Cette modification est confirmée par une diminution de l'activité de la thioestérase lors de l'utilisation d'inhibiteurs de protéases à cystéine. De plus, la substitution de la cystéine par une sérine a permis de retrouver une sensibilité aux inhibiteurs à sérine et une diminution de l'activité (Yuan *et al.*, 1995). La conformation particulière de la thioestérase chez *Arabidopsis thaliana* a permis de définir les thioestérases à triade catalytique « papain like » (Turk *et al.*, 1997; Cantu *et al.*, 2010).

1.3.2. Les enzymes lipolytiques animales

Le domaine des enzymes lipolytiques animales étant vaste, ce sous-chapitre développera les études les plus importantes. Dans un premier temps, les lipases seront présentées. Les phospholipases, carboxylestérases et thioestérases seront traitées dans un second temps. Enfin, de par leurs particularités, les enzymes lipolytiques d'animaux marins et d'insectes seront présentées dans la dernière partie.

Lipases

Les graisses représentent chez les animaux une source énergétique importante. Les voies métaboliques utilisant cette source énergétique sont donc essentielles pour l'équilibre physiologique des organismes. Le modèle enzymatique de référence pour le règne animal est la lipase pancréatique (Hofmann and Borgstrom, 1963). Cependant, d'autres enzymes ont été largement étudiées comme la lipase gastrique et la lipase hépatique. La structure des gènes codant pour ces enzymes laisse supposer un ancêtre commun (Warden *et al.*, 1993; Oku *et al.*, 2006). L'idée d'une identification et d'un classement des gènes de lipase selon leur composition a alors été émise (Wong and Schotz, 2002). Si au départ, cette famille de gènes ne comporte que trois lipases : lipoprotéine-lipase, lipase pancréatique et lipase hépatique, l'outil de comparaison de séquences a rapidement permis d'en

découvrir d'autres par homologie de leur séquence protéique. De nouvelles enzymes ont trouvé leur place dans la famille des lipases comme, par exemple, la lipase endothéliale et la phosphatidylsérine phospholipase A1. La plupart des membres de la famille des lipases ont été étudiés d'un point de vue physiologique ou pour établir leur distribution à travers les tissus. Un type de lipase, la lipase hormono-sensible (HSL) a au début été décrite comme une lipase spécifique des triglycérides et intracellulaire aux adipocytes. Actuellement, elle a été identifiée dans la plupart des tissus comme les lipoprotéines lipases. La principale caractéristique de la lipase hormono-sensible (HSL) reste la variété de ses régulations notamment par phosphorylation/déphosphorylation ou par interaction avec d'autres protéines (Yeaman, 2004).

Les processus de dégradation des graisses sont réalisés par différentes enzymes. La lipase gastrique dégrade les lipides des aliments au niveau de l'estomac. C'est une lipase extrêmement résistante aux conditions d'acidité (pH proche de 1) et aux attaques notamment de la pepsine et des protéases au niveau de l'estomac (Canaan *et al.*, 1999). La lipase pancréatique est sécrétée par les cellules acineuses au niveau du duodenum. Elle a été largement étudiée chez de nombreux organismes. Elle est active à pH légèrement alcalin par opposition à la lipase gastrique. La lipase pancréatique humaine a fait l'objet de nombreuses études permettant de mieux comprendre sa régulation et son fonctionnement. L'activité de la lipase pancréatique est inhibée aux concentrations physiologiques en sels biliaires (Verger, 1984). Cette activité est restaurée en présence d'une autre protéine pancréatique de 11 kDa appelée colipase. Il s'agit d'un excellent exemple d'interaction protéine-protéine. En présence de bile, la lipase et la colipase forment en effet un complexe qui se positionne à l'interface lipide/eau, où il peut être adsorbé. La lipase pancréatique bascule alors en conformation active (Figure 11). La libération des acides gras peut provoquer l'arrêt de l'activité s'ils ne sont pas pris en charge par les sels biliaires (Borgstrom and Erlanson, 1973). Les sels biliaires permettent également le maintien de l'émulsion et donc de l'interface sans laquelle l'activité serait nettement réduite (Lowe, 1997, 1999).

Phospholipases, carboxylestérases et thioestérases

Les phospholipases au sein du règne animal forment une très vaste famille divisée comme vu précédemment en fonction de leur site d'action (Figure 5). Pour chaque sous-famille de phospholipases (A, C et D), des caractéristiques physiologiques et spécifiques ont été mises en évidence par rapport aux tissus. La phospholipase A1 compte de nombreux isozymes comme la lipase hépatique et la lipase pancréatique capables d'hydrolyser des triglycérides mais également des phospholipides. Les phospholipases A possèdent le pentapeptide GX SXG (elles appartiennent donc à la famille génétique des lipases). Elles sont classées en fonction de leur localisation de part et d'autre de la paroi cellulaire. Les phospholipases A2 se répartissent en cinq groupes : les PLA2 sécrétées, les PLA2 calcium dépendantes, les PLA2 calcium indépendantes, les PLA2 lysosomales et enfin les PLA2 acetylhydrolases (PAF-AH). Ces phospholipases A2 sont très largement distribuées à travers les tissus et leurs structures tridimensionnelles sont bien distinctes en fonction des groupes de PLA2. Les phospholipases C représentent au sein du règne animal 13 isozymes classés en six groupes (β , γ , δ , ϵ , ζ et η). Les phospholipases β et ϵ sont activées par des récepteurs couplés à des protéines G. La distribution au sein des tissus est variable mais la représentation des phospholipases C au sein des tissus du cerveau est conservée pour 5 des 6 groupes. L'activité phospholipase C au niveau de ces cellules conditionne la bonne modulation des réseaux neuronaux. Comme pour les phospholipases A2, chacun des six groupes adopte une conformation structurale propre. La figure 13 présente la schématisation des domaines constitutifs pour chacun des groupes, certains domaines sont conservés comme les domaines catalytiques « X » et « Y » ou bien le domaine de liaison au calcium « C2 » (Bamji-Mirza and Yao, 2011).



Figure 13: Schématisation des structures d'isozymes de phospholipase C chez la souris.

PH (domaine pleckstrin homology), EF (motif « main EF »), X (domaine catalytique X), Y (domaine catalytique Y), C2 (domaine de liaison du Ca^{2+}), SH (domaine d'homologie Src), Ras GEF (facteur d'échange nucléotidique Ras-guanine), chaque nombre indique la longueur en acides aminés de chaque isozyme.

Les phospholipases D sont représentées par seulement deux isozymes, D1 et D2. De par leur rôle important dans la mise à disposition notamment de l'acide phosphatidique, les phospholipases D sont sous le contrôle de nombreux co-facteurs. Parmi ces co-facteurs, la GTPase ou l'actine F

induisent l'activation de l'enzyme, l'actine G active son inhibition. Les phospholipases D sont également sous le contrôle d'ions divalents comme le calcium (Selvy *et al.*, 2011).

Une autre famille de phospholipase vient compléter celles précédemment décrites, c'est la famille des phospholipases à domaine patatine. Le domaine patatine a été décrit en premier chez les plantes et se caractérise par un site catalytique ne comportant que deux acides aminés Ser et Asp. Différentes enzymes homologues ont été mises en évidence chez l'homme. Leur distribution est large dans les tissus (tissus adipeux, cœur, système nerveux) et la variété de séquences a permis l'élaboration d'une famille spécifique de ce type d'enzyme (Wilson *et al.*, 2006).

Pendant de nombreuses années, l'hydrolyse de lipides a été associée aux lipases. La réalisation de ce type d'hydrolyse par des carboxylestérases a été mis en évidence plus tard par l'observation d'activités lipolytiques indépendantes des lipases (Soni *et al.*, 2004). Les carboxylestérases pouvant participer à l'activité lipolytique sont présentes majoritairement au niveau du foie et de l'intestin grêle chez la plupart des mammifères (Satoh and Hosokawa, 2006). Les carboxylestérases de mammifères ont dans un premier temps été classées dans la vaste superfamille des estérases en fonction de la spécificité de leurs substrats. Récemment, une nouvelle classification tenant compte des degrés de similarité et de l'homologie entre séquences a permis de définir une famille multigénique dont les isozymes sont réparties en cinq groupes (CES1, 2, 3, 4A et 5A). Leur rôle est la dégradation de composés chimiques comportant au moins une chaîne d'acyle afin de la libérer. Ces dégradations ont lieu notamment dans les processus de détoxification cellulaire. La troisième fonction possible pour les carboxylestérases humaines est l'hydrolyse de composés endogènes comme les triglycérides à courtes ou longues chaînes mais également des acyl-coenzymes A. Ces enzymes ont ainsi été mises en évidence, car elles complétaient l'hydrolyse des triglycérides réalisée par les lipases (Taketani *et al.*, 2007; Satoh and Hosokawa, 2010).

Les thioestérases animales ont montré des activités lipolytiques sur des acides gras à moyenne et longue chaîne. Leur classification a été établie en fonction de leur conformation et des résidus intervenant dans le site catalytique. Ces thioestérases sont classées parmi les α/β hydrolases, elles adoptent des conformations de carboxylestérase (Cantu *et al.*, 2010).

Au sein du règne animal, les activités lipolytiques peuvent donc être réalisées par de nombreuses enzymes. Une étude de 2005 sur le tissu adipeux de souris a montré la présence de 23 protéines à activité lipolytique non identifiées comme des lipases (Birner-Gruenberger *et al.*, 2005).

Les enzymes lipolytiques animales d'origine marine

Les organes de poissons sont connus pour contenir nombre de molécules d'intérêt, notamment des acides gras de la série $\omega 3$. La présence de ces acides gras laisse supposer la présence d'enzymes lipolytiques (lipases et phospholipases) pouvant se révéler indispensables pour l'enrichissement d'aliments en acides gras poly-insaturés issus de sous-produits de la pêche (Shahidi and Janak Kamil, 2001). Ces acides gras sont en effet plus difficiles à libérer par hydrolyse avec les lipases actuellement utilisées (Lie and Lambertsen, 1986). Les lipases d'organismes marins se sont à l'inverse montrées plus efficace pour ces acides gras à longue chaîne (Iijima *et al.*, 1998). Il faut donc approfondir les connaissances sur ces possibles activités lipolytiques présentes dans les différents tissus de produits de la mer (Smichi *et al.*, 2012).

Comme pour toutes les espèces, les lipides sont des acteurs importants du métabolisme. Cette notion s'applique également aux espèces marines, chaque espèce va adapter son métabolisme en fonction de son régime alimentaire. Certains animaux marins vont donc orienter leur métabolisme vers les graisses alors que d'autres vont l'orienter vers les sucres (les enzymes lipolytiques seront donc moins sollicitées). La voie métabolique empruntée par chaque type d'organisme va conditionner la taille et l'activité de certains organes. Chez certaines espèces, des particularités physiologiques (emplacement de glandes digestives notamment) vont apparaître. Une étude récente a montré la présence d'activité lipolytique dans les glandes digestives au sein d'organismes marins de différents groupes (Smichi *et al.*, 2012). De plus, chez les animaux marins comme chez les animaux terrestres, les activités lipases sont plus importantes que les activités phospholipases (Borgström and Brockman, 1984). Ceci peut être mis en corrélation avec les faibles concentrations en phospholipides et les plus importantes concentrations en triglycérides dans les nutriments. A travers les groupes d'animaux marins, l'activité lipolytique varie. Les poissons cartilagineux montrent une activité lipolytique plus importante que les poissons à squelette osseux, les mollusques ou bien encore les crustacés. Cette différence d'activité peut être due au régime alimentaire. En effet, les poissons cartilagineux sont majoritairement carnivores alors que les poissons osseux eux le sont moins. Quant aux mollusques et crustacés, ils sont pour leur part peu consommateurs de nutriments riches en acides gras. Les caractéristiques des enzymes lipolytiques sont dépendantes des espèces, notamment pour les valeurs de température et de pH optimales. Concernant les colipases, co-facteurs parfois nécessaires à l'activité lipolytique, elles sont présentes chez les mammifères et les oiseaux et semblent présentes chez les poissons cartilagineux. Elles seraient néanmoins absentes chez les mollusques, crustacées et les autres poissons (Sternby *et al.*, 1983; Smichi *et al.*, 2012).

Les enzymes lipolytiques d'insectes

Les enzymes lipolytiques d'insectes ont été caractérisées pour la 1^{ère} fois lors de l'hydrolyse de lipides de stockage au stade larvaire chez la mouche, *Calliphora erythrocephala* (Price and Russell, 1975). Chez les insectes, de nombreuses enzymes lipolytiques ont été identifiées et localisées dans différents tissus. Les lipides de stockage représentent une des deux principales sources d'énergie avec le glycogène. Ces lipides sont localisés au niveau de cellules adipeuses et principalement au niveau d'une structure de réserve appelée « corps gras » (Arrese and Soulages, 2010). Le corps gras réalise la majorité des synthèses de protéines de l'hémolymph, son rôle peut être assimilé à celui du foie et des tissus adipeux chez les animaux supérieurs (Arrese *et al.*, 2006). La mobilisation de ces réserves lipidiques peut intervenir pour l'apport d'énergie aux muscles, pour le développement des œufs, et lors des périodes de non-alimentation.

L'activité lipolytique et la présence de lipases évoluent en fonction du stade de développement. Elle est soumise à une régulation hormonale par un neuropeptide. Celui-ci semble intervenir autant pour la mobilisation des triglycérides que pour la mobilisation des réserves sous forme de sucre. Des récepteurs pour cette hormone ont été identifiés chez la drosophile et le ver à soie (Staubli *et al.*, 2002; Gade and Auerswald, 2003). Différentes enzymes homologues de lipases ont été caractérisées chez la drosophile, Lip1, Lip2 et Lip3. Elles possèdent le motif GHSQG correspondant au motif consensus prédéfini (GXSXG). Lip1 et Lip3 sont exprimées dans différents tissus et à des stades de développement différents, laissant supposer des rôles bien spécifiques. La protéine Lip1 correspond, par sa séquence, à des lipases digestives et est retrouvée au stade larvaire précoce. Le gène Lip3 semble être exprimé à un stade larvaire plus avancé, le gène Lip2 semble lui non fonctionnel. Une autre étude a mis en évidence la présence d'une lipase au niveau de la glande accessoire de la drosophile. Cette lipase a été révélée par une activité lipolytique résiduelle en l'absence de la carboxylestérase « Est6 » généralement présente dans ce tissu (Smith *et al.*, 1994; Pistillo *et al.*, 1998).

1.3.3. Les enzymes lipolytiques de micro-organismes

Ces enzymes sont les plus répandues chez les bactéries, les levures et les champignons filamenteux (Treichel *et al.*, 2010). De nombreuses bases de données permettent de répertorier l'ensemble des enzymes lipolytiques trouvées dans le monde microbien. Une classification a été proposée à partir de deux critères : la séquence protéique et les propriétés biologiques fondamentales (Arpigny and Jaeger, 1999).

Les bactéries

Les bactéries produisent la plupart des classes d'enzymes lipolytiques : des enzymes classées comme lipases ou carboxylestérases jusqu'aux phospholipases. La classification d'Arpigny and Jaeger (1999) a servi de base pour définir huit familles et y inclure les différentes enzymes hormis les phospholipases. La base de cette classification est structurale, les enzymes doivent appartenir à la classification des α/β hydrolases décrites dans le paragraphe 1.2.1.

La famille I (comprenant six sous-familles) adopte le classement décrit auparavant par Jaeger *et al.* (1994) et se consacre aux “ lipases vraies ”. Les lipases issues du genre *Pseudomonas sp.* sont les premières décrites et inventoriées. Cette première famille inclut également les lipases de *Bacillus subtilis* ou d'autres bactéries à Gram positives. Pour deux sous-familles I¹ et I², l'expression fonctionnelle d'une lipase est dépendante de la présence d'une protéine chaperonne appelée foldase assurant la bonne conformation de l'enzyme.

La famille II comprend les enzymes lipolytiques ne portant pas le classique pentapeptide GX SXG mais un motif GDS(L) contenant la sérine du site actif. Une autre particularité de cette famille est la proximité du site actif par rapport à la partie N-terminal de la protéine (Upton and Buckley, 1995; Akoh *et al.*, 2004).

Les enzymes de la famille II peuvent agir sur une grande variété de substrats. Cette propriété est due à un site catalytique modulable correspondant au modèle de Koshland. L'hypothèse est la possible adaptation du site catalytique d'une enzyme à la nature du substrat à dégrader (Koshland, 1958). Les activités isolées des enzymes de la famille II peuvent être du type thioestérase, protéase arylestérase ou bien encore lysophospholipase. L'étude approfondie des enzymes composant cette famille a conduit à proposer une sous-classification de manière à exposer au mieux les variations de structure et de composition des enzymes GDS(L). Cette classification tient notamment compte de différents blocs peptidiques conservés au sein des séquences protéiques (Akoh *et al.*, 2004).

Les enzymes de la famille III ont une conformation d'hydrolase α/β avec un site consensus classique. Ces enzymes présentent par comparaison de séquence une identité de 20 % avec les phospholipases A2 AcetylHydrolase PAF-AH (vu au paragraphe 1.3.2). Cette famille d'enzymes est également caractérisée par l'homologie des séquences avec la lipase isolée de *Streptomyces exfoliatus* (30 à 80 % d'identité).

La famille IV regroupe des enzymes disposant d'une similarité avec les lipases hormono-sensibles (HSL) des mammifères, notamment pour le domaine catalytique. Un modèle a donc été proposé pour définir les acides-aminés du site actif. Chez les mammifères, un domaine existe en position N-terminal ainsi qu'un module de régulation supplémentaire, montrant une évolution par rapport aux séquences bactériennes. Depuis la classification de 1999, cette famille a pu être plus étudiée. Il apparaît ainsi, qu'en dépit de la présence de similarités avec les lipases HSL de mammifères, les activités des lipases de la famille IV demeurent restreintes aux chaînes courtes d'acides gras, alors que pour le cas des enzymes de mammifères, l'activité concerne également les esters insolubles (Chahinian *et al.*, 2005).

Les enzymes de la famille V appartiennent comme pour la famille IV à des organismes mésophiles (*Pseudomonas oleovorans*, *Haemophilus influenza*, *Acetobacter pasteurianus*). Ces enzymes partagent des similarités de séquence (20-25 %) avec des enzymes non-lipolytiques de la famille des hydrolases α/β . Cette famille a été élargie depuis à des enzymes issues d'organismes adaptés à des hautes températures comme *Helicobacter pylori*. Un gène codant une enzyme lipolytique (EstV) issue de cette bactérie a notamment été cloné. Il a été caractérisé comme comportant de nombreuses similitudes avec des lipases, néanmoins il dispose d'une cinétique enzymatique Michaëlienne (Ruiz *et al.*, 2007).

La famille VI contient des enzymes de faible poids moléculaire (23 à 26 kDa) identifiées comme les plus petites carboxylestérases connues. Une carboxylestérase de *Pseudomonas aeruginosa* dont la structure a été établie en 2005 permet aujourd'hui d'orienter la fonction des enzymes de cette famille. En effet, des similarités de structures et de séquences ont notamment été caractérisées avec la thioestérase acyl protéine 1 humaine (Pesaresi and Lamba, 2005).

La famille VII comporte des carboxylestérases plus communes partageant des similarités de séquence avec des enzymes eucaryotes (acétylcholine estérase et carboxylestérase hépatique).

La famille VIII est composée d'enzymes d'environ 380 acides aminés similaires aux bêta-lactamases de classe C. Cette dernière famille comprend également des enzymes contenant non seulement le motif GX SXG mais également le motif SXXL.

Les approches bio-informatiques, l'augmentation des données génomiques et le développement d'outils de clonage et d'expression permettent la caractérisation de nombreuses enzymes. De nouvelles enzymes ont ainsi été identifiées et ont permis une mise à jour de la classification de 1999 (Hausmann and Jaeger, 2010). Cette classification a également montré la présence d'homologie avec

des enzymes non lipolytiques et la nécessité d'une caractérisation biochimique d'une enzyme (Ali *et al.*, 2012). La comparaison de séquences reste un outil important pour la prédiction de la fonction protéique. De nouvelles séquences et de nouveaux motifs sont apparus, comme par exemple la séquence consensus GXXSXXG, mais l'activité lipolytique n'a pas encore été démontrée à ce jour.

Les phospholipases sont également présentes chez les bactéries. Les phospholipases A, localisées sur la membrane externe des bactéries Gram négatives sont, par exemple, impliquées dans l'attaque de la membrane des cellules hôtes à infecter (Istivan and Coloe, 2006). Les phospholipases C sont également impliquées dans ce type d'activité. Certaines phospholipases C hydrolysent la phosphatidylcholine en phosphorylcholine et en diacylglycérol. D'autres dégradent des composés comme le phosphatidylinositol ou la sphingomyéline (Korbsrisate *et al.*, 2007). La production de diacylglycerol et de produits de dégradation comme l'acide arachidonique peuvent engendrer des cascades catalytiques modifiant les mécanismes de régulation et provoquant des inflammations cellulaires (Titball, 1993). Les phospholipases D, sont responsables de la virulence de certaines bactéries comme *Corynebacterium pseudotuberculosis* ou *Acinetobacter baumannii* (Hodgson *et al.*, 1990; McKean *et al.*, 2007; Jacobs *et al.*, 2010).

Les champignons filamenteux et les levures

Les lipases commerciales les plus répandues sont issues des espèces de champignons filamenteux tels que *Rhizopus sp.*, *Aspergillus sp.*, *Penicillium sp.*, *Mucor sp.* et *Rhizomucor sp.*. La capacité de production varie d'une espèce à l'autre en fonction de la composition du milieu de culture, de la source d'azote et de carbone et en fonction du pH et de la température (Cihangir and Sarikaya, 2004). Parmi les enzymes caractérisées chez les champignons filamenteux, la famille des cutinases est importante. Ces enzymes sont responsables d'une partie du rôle pathogène de certains champignons. Elles représentent un fort potentiel d'utilisation (Sorensen *et al.*, 2007; Gaillardin, 2010).

Plusieurs espèces de levures produisent des lipases d'intérêt. *Candida rugosa* est la plus fréquemment utilisée pour obtenir des lipases (Vakhlu and Kour, 2006). Ce microorganisme est présent dans l'eau, les produits laitiers et le tube digestif de l'homme. Différentes espèces de *Candida* ont été particulièrement étudiées afin d'obtenir différentes enzymes et fonctionnalités. Il faut notamment retenir *Candida antarctica* et *Candida cylindracea* (Lotti *et al.*, 1993; Pandey *et al.*, 1999). La plupart des enzymes lipolytiques issues de levures sont extracellulaires, à l'exception de *Candida parapsilosis* qui produit une lipase intracellulaire et de *Yarrowia lipolytica* qui produit des lipases à la fois intra- et extracellulaires. Une majorité de levures synthétisent au moins une lipase, et jusqu'à une dizaine

d'isozymes ont pu être identifiés. Les enzymes obtenues étant extracellulaires, le procédé de production peut être réalisé par fermentation à l'état solide. De nombreuses études ont montré l'importance à la fois de la technique de croissance des microorganismes producteurs, mais également du milieu de culture. Pour *Candida*, Chang *et al.* (1994) ont montré que non seulement la méthode de culture pouvait influencer la production de lipases, mais également agir sur la conformation de l'enzyme (production de différentes isoformes). La composition du milieu de culture influencerait la régulation et l'expression du gène codant la lipase mais serait également impliqué dans les modifications post-traductionnelles. Ces différentes isoformes peuvent présenter des sélectivités, des activités spécifiques, et des sensibilités aux inhibiteurs très différentes. L'homologie de séquence semble drastiquement conservée entre les différentes isozymes, mis à part pour la zone d'interaction entre l'enzyme et le substrat (Lotti *et al.*, 1994).

D'autres espèces que *Candida* sont connues pour produire des lipases. Le genre *Geotrichum* produit des lipases essentiellement extracellulaires (Tsujioka *et al.*, 1973). *Geotrichum candidum* est une espèce assez répandue dans les sols, l'eau, l'air, les céréales et dans les produits laitiers (affinage de certains fromages). Comme pour *Candida*, la levure *Geotrichum candidum* présente de nombreuses isoformes d'enzymes lipolytiques. Les caractéristiques biochimiques et les spécificités vis-à-vis du substrat dépendent également du milieu de production de l'enzyme (Stránský *et al.*, 2007).

La levure *Trichosporon* a aussi été largement étudiée pour la production de lipase. Une des espèces, *Trichosporon fermentans*, produit une lipase dont la quantité est fonction de la source de carbone disponible dans le milieu. Avec une source de carbone comme l'huile d'olive, la production de lipase est quatre fois plus importante qu'avec le glucose (Chen *et al.*, 1992). Cette constatation a depuis été utilisée pour optimiser la production de lipases dans d'autres organismes (Wang *et al.*, 2008).

Les quatre types de phospholipases sont présents chez les microorganismes. Elles sont impliquées, le plus souvent, dans des processus de modifications membranaires ou de synthèse de molécules bioactives. Certaines de ces phospholipases sont des éléments clefs dans la virulence ou le rôle pathogène de leur cellule productrice. C'est le cas notamment des phospholipases des levures pathogènes du type *Candida* sp.. Ibrahim *et al.* (1995) ont montré que l'infection par *Candida albicans* induisait la mortalité de lignées cellulaires de souris. Ils ont notamment caractérisé la présence d'une grande quantité de phospholipases extracellulaires. Dans les champignons, de nombreuses membranes cellulaires sont constituées de phospholipides. Ces organismes peuvent moduler la composition de leur membrane par action des phospholipases. Chez *Saccharomyces*

cerevisiae, des phospholipases D sont par exemple impliquées dans la formation de la membrane lors de la méiose ou lors de la sporulation. C'est également le cas lors de la transition dimorphique chez *Candida albicans* (Rose *et al.*, 1995; Rudge *et al.*, 1998; Dolan *et al.*, 2004). Les phospholipases A2 et B sont retrouvées dans de nombreuses espèces de champignons dont certaines espèces pathogènes opportunistes comme *Candida albicans*. Au sein de ces organismes, les fonctions des phospholipases peuvent autant être physiologiques dans l'adaptation cellulaire que pathogènes lors d'infections fongiques (Kohler *et al.*, 2006).

L'activité thioestérase est également rencontrée chez les champignons et les levures. Ces organismes ont notamment présenté les activités Acyl-CoA hydrolase, Acyl-ACP hydrolase, protein-acyl hydrolase et glutathione hydrolase (Jones *et al.*, 1999; Ntamack *et al.*, 2009).

1.4. Les enzymes lipolytiques de microalgues

La majorité des microalgues est dite photo-autotrophe ou autotrophe. Elles ont la capacité de transformer l'énergie lumineuse en énergie sous forme d'ATP, permettant le stockage du CO₂ sous forme de lipides et d'hydrates de carbone. De par cette caractéristique, les microalgues peuvent être considérées comme organismes oléagineux (National Renewable Energy Laboratory, NREL, 1998). Les lipides peuvent représenter jusqu'à 86 % du poids total sec des microalgues. Cette proportion est physiologique et représente généralement une réponse à des changements environnementaux (Murphy, 2001). La présence globale de fortes teneurs en lipides laisse supposer la présence d'enzymes lipolytiques innovantes et spécifiques au sein des microalgues (Godet *et al.*, 2010). A ce jour, très peu d'enzymes lipolytiques ont été caractérisées chez les microalgues. Ce sous-chapitre traitera dans un premier temps des phospholipases puis des autres activités lipolytiques détectées chez les microalgues. Une dernière partie sera consacrée à la microalgue *Isochrysis galbana*, modèle biologique utilisé au laboratoire.

Les phospholipases

Des enzymes lipolytiques d'origine algale ont été mises en évidence. La plupart sont des phospholipases. Bilinski *et al.* (1968) ont, par exemple, caractérisé une phospholipase C chez la microalgue *Monochrysis lutheri*. L'activité de l'enzyme a été mesurée, à partir d'extraits protéiques totaux, par hydrolyse de la phosphatidylcholine marquée au ¹⁴C. La purification de cette phospholipase C n'a cependant pu être réalisée par manque de matériel biologique.

Peu après, en 1970, une phospholipase D a été identifiée et caractérisée chez la microalgue rouge *Porphyridium cruentum* (aussi connue sous le nom *Porphyridium purpureum*). L'activité enzymatique a été estimée à partir de lysats cellulaires obtenus par sonication (Antia *et al.*, 1970). Une libération de choline montre une activité phospholipase D sur de la lécithine marquée au ^{14}C . Cette microalgue est actuellement commercialisée et utilisée pour ces apports en acides gras polyinsaturés $\omega 3$ (eicosapentaénoïque, EPA) et $\omega 6$ (eicosatétraénoïque, ARA).

Chez la diatomée *Thalassiosira rotula*, une phospholipase A2 a été détectée. Cette enzyme est à l'origine d'une cascade de réactions aboutissant à la libération d'acides gras. Ce phénomène est déclenché comme réaction de défense cellulaire chimique par la microalgue (Pohnert, 2002).

La chlorophycée *Chlamydomonas reinhardtii* apparaît comme étant la microalgue la plus étudiée, avec 7 000 articles et 150 brevets. L'assemblage de son génome n'est cependant pas encore complet. L'activité lipolytique chez cette microalgue a récemment été mise en évidence, notamment une activité phospholipase. Ce modèle a depuis longtemps été choisi pour caractériser des réponses à des stress en milieu marin, notamment des stress de salinité. De nombreux processus entrent en jeu pour répondre à ce type de stress. L'un d'eux est une cascade déclenchée par une activité phospholipase A2 liée à une activité diacylglycérol kinase. Cette activité génère la libération d'acide lysophosphatidique (Arisz and Munnik, 2011). Des activités lipolytiques ont aussi été observées chez *Dunaliella salina* (Roessler, 1990). Elles ont été en partie caractérisées dans les membranes des organelles et une a été identifiée comme activité phospholipase C.

Autres activités lipolytiques

Parmi les autres activités lipolytiques, une activité lipase ciblant les galactoglycérolipides a été mise en évidence en 2012 chez *C. reinhardtii*. Cette activité très spécifique permet de réguler et d'organiser la mobilisation des réserves d'énergie lors de la défense de la cellule en cas de stress. L'utilisation de mutants dépourvus du gène codant cette lipase, a abouti dans les mêmes conditions de stress, à une perte de viabilité notable (Li *et al.*, 2012a; Li *et al.*, 2012b). Le métabolisme des triglycérides a été étudié chez *C. reinhardtii*. Un gène de lipase, CrLIP1, a été identifié. Il est impliqué dans l'hydrolyse des diacylglycérols lors du retour à un taux normal d'azote dans l'environnement (Li *et al.*, 2012a). L'activité acyltransférase a également été mise en évidence pour la production de triglycérides. L'enzyme concernée semble de plus, réaliser l'hydrolyse d'acyles notamment sur des substrats de type galactolipides, phospholipides et triglycérides (Yoon *et al.*, 2012). Dans une autre étude chez *Dunaliella salina*, Einspahr *et al.* (1989) ont observé une activité phosphomonoestérase. Pour la

microalgue *Skeletonema costatum*, la présence d'une probable activité glycolipase a été supposée (Bergé JP, Thèse 1996). Enfin en 2010, Treichel *et al.* (2010) ont purifié et caractérisé une lipase de la cyanobactérie *Spirulina platensis*.

De manière générale, les microalgues réagissent à des modifications de leur environnement par une modification de leur composition en acides gras (Thompson *et al.*, 1992).

La microalgue la plus abondante du phytoplancton, *Emiliana huxleyi* est une des premières souches de microalgue dont le génome a été entièrement séquencé (Sanchez Puerta *et al.*, 2005). Néanmoins, à ce jour, aucune activité lipolytique n'a été caractérisée chez cette microalgue coccolithophore. Le génome de plusieurs microalgues a été séquencé mais, mis à part pour *Pheodactylum tricornutum*, aucun gène présentant une activité lipolytique bien caractérisée n'a été cloné. Chez cette microalgue, un gène de thioestérase a été isolé et caractérisé (Gong *et al.*, 2011).

Isochrysis galbana

Chez la microalgue *I. galbana*, une activité lipolytique endogène a été suspectée suite à la présence d'acides gras libres dans les lipides totaux (Poisson L, Thèse 1999). En 2005, la mise en évidence d'activités glycolipasiques et phospholipasiques intracellulaires ont conforté cette hypothèse (Devos M, Thèse 2005).

En effectuant une recherche bibliographique via la base de données PubMed (source: <http://www.ncbi.nlm.nih.gov/pubmed>) avec des mots clés associant le terme “ *Isochrysis galbana* ” et les termes “ carboxylestérase, lipase ou estérase ou enzyme lipolytique ”, les résultats sont peu nombreux comparativement à d'autres espèces. L'analyse bioinformatique de séquences annotées a permis d'envisager la présence d'enzymes lipolytiques qui pourraient intervenir dans le métabolisme lipidique chez *I. galbana* (Godet *et al.*, 2010).

1.5. Les applications industrielles des enzymes lipolytiques

Les enzymes sont des catalyseurs naturels provenant de divers organismes vivants. Elles sont responsables d'une variété et d'un nombre important de réactions. Par leur évidente implication dans des processus moléculaires au niveau de la cellule, les enzymes occupent une place importante dans le monde vivant. Les enzymes sont ainsi devenues des outils indispensables au monde industriel. La nature des enzymes utilisées dans l'industrie est variable, mais toutefois le marché est dominé par les enzymes à activité hydrolytique. Cette proportion représente 75 % du marché des enzymes et inclut des carbohydrases, des protéases et plus récemment des lipases (Li *et al.*, 2012a). Les lipases sont placées derrière les protéases et les carbohydrases dans le classement mondial des enzymes commerciales et représentent actuellement 5 % de ce marché. La distribution de ces enzymes est inégale dans les applications industrielles. Ce sont, en effet, les enzymes issues de microorganismes qui, étant plus stables aux températures et pH industriels, sont préférées aux enzymes d'autres origines (Dutra *et al.*, 2008; Griebeler *et al.*, 2011). Elles ont comme avantage d'être produites facilement et d'avoir des caractéristiques de résistance et d'activité dans des milieux et des conditions extrêmes. De plus, les lipases d'origine microbienne ne nécessitent pas forcément de co-facteurs pour être actives et peuvent agir sur une large gamme de substrat. Ces enzymes peuvent être actives sous forme immobilisée ou en solution (de Azeredo *et al.*, 2007). Les lipases sont retrouvées dans de nombreuses applications comme par exemple dans le recyclage ou la fabrication de papier mais également dans la décomposition de matière grasse sur des tissus en étant associées à des détergents.

Ces enzymes représentaient en 2010, un marché de 1 milliard de dollars et devraient représenter environ 1,5 milliard de dollars en 2015 (Enzymes in industrial applications : Global markets, 2011). Depuis quelques années, l'importance des enzymes lipolytiques dans les processus industriels a significativement augmenté. Les fournisseurs dominants de ce secteur sont Novozymes®, Amano-Enzyme, DuPont® ou encore DSM. Cette tendance et la demande qui en résulte définissent actuellement deux stratégies : la recherche de nouvelles enzymes et l'optimisation d'enzymes existantes.

- La recherche de nouvelles enzymes est réalisée notamment dans des milieux extrêmes ou spécifiques (pH, salinité, milieux de culture non reproductibles en laboratoire). La métagénomique permet ainsi de caractériser des activités lipolytiques directement depuis des échantillons de sols, de marais, ou de zone de balancement des marées. L'isolement de l'organisme à l'origine de l'activité n'est plus nécessaire (Zhang, 2009; Gilbert *et al.*, 2011).

- L'optimisation et la modification d'enzymes déjà identifiées sont réalisées notamment par mutagenèse dirigée ou aléatoire. C'est le cas, par exemple, de la lipase A de *Candida antarctica* dont le site catalytique a été modifié afin de spécialiser son action. L'activité principale de la lipase A est alors devenue peu active sur les esters 2-phenylpropanoate mais très active sur l'ester d'intérêt 2-(4-isobutylphenyl)propanoate (ibuprofen) (Sandstrom *et al.*, 2012).

1.5.1. Méthodes de production des lipases

Les études sur les lipases présentant un intérêt industriel rapportent différentes méthodes de production en fonction de l'origine de la lipase. Les lipases issues d'espèces de champignons sont plus favorablement produites par culture en milieu solide alors que celles issues d'organismes comme les levures ou les bactéries sont plus souvent produites par culture submergée (Dutra *et al.*, 2008). Même si certaines études continuent de décrire des productions d'enzymes dans des Erlenmeyers, il est aujourd'hui courant d'utiliser des bioréacteurs. Le tableau 2 résume les différentes formes de production de lipases ainsi que les configurations des bioréacteurs (Treichel *et al.*, 2010).

	Type de procédure	Type de bioréacteur	Références
Culture submergée	En discontinu (batch)	Agitation mécanique	Gupta <i>et al.</i> , 2007; Burkert <i>et al.</i> , 2005; Montesinos <i>et al.</i> , 1997; Potumarthi <i>et al.</i> , 2008; Cos <i>et al.</i> , 2005; Alonso <i>et al.</i> , 2005; Gordillo <i>et al.</i> , 1998; Kar <i>et al.</i> , 2008; Becker and Markl, 2000; D'Annibale <i>et al.</i> , 2006; Puthli <i>et al.</i> , 2006; Fickers <i>et al.</i> , 2006; Zhao <i>et al.</i> , 2008
	En discontinu (batch)	Air pulsé	Burkert <i>et al.</i> , 2005
	En discontinu (batch)	Collonne à bulle	D'Annibale <i>et al.</i> , 2006; Ghadge <i>et al.</i> , 2005
	En discontinu (répété)	Agitation mécanique	Benjamin and Pandley, 1997; Li <i>et al.</i> , 2005; Yang <i>et al.</i> , 2005
	En batch alimenté	Agitation mécanique	Boareto <i>et al.</i> , 2007; Montesinos <i>et al.</i> , 1997; Gordillo <i>et al.</i> , 1998; Cos <i>et al.</i> , 2005; Gordillo <i>et al.</i> , 1998; Ito <i>et al.</i> , 2001; Surribas <i>et al.</i> , 2007; Kim and Hou, 2006; Ikeda <i>et al.</i> , 2004
	En continu	Agitation mécanique	Montesinos <i>et al.</i> , 1997,2003; Jensen <i>et al.</i> , 2002; Gordillo <i>et al.</i> , 1998; Becker and Markl, 2000
Culture en milieu solide	En discontinu	Lit fixe	Cavaleanti <i>et al.</i> , 2005; Benjamin and Pandey, 2001; Diaz <i>et al.</i> , 2006; Martinez <i>et al.</i> , 2008
	En discontinu	Plateau	Mala <i>et al.</i> , 2007

Tableau 2: Présentation des principales méthodes de production de lipases

1.5.2. Applications dans les détergents

La majeure partie des enzymes lipolytiques obtenues à l'échelle industrielle est utilisée pour la production de détergents. Les lipases sont ajoutées aux détergents à usage domestique ou industriel. Elles sont présentes dans les détergents de lave-vaisselles ainsi que pour le débouchage des canalisations. Leurs actions sont évidemment ciblées sur les résidus gras ; elles sont utilisées en association avec d'autres enzymes comme les amylases ou les cellulases (Ito *et al.*, 1998). L'utilisation accrue de détergents depuis l'apparition des lave-linges a eu un impact environnemental. Pour diminuer cet impact et maintenir l'efficacité de lavage, les méthodologies se sont également orientées vers l'ajout d'enzymes aux détergents. L'objectif étant de réduire l'usage d'agents chimiques

difficilement éliminables et ainsi d'obtenir des produits plus respectueux de l'environnement (Gavrilescu and Chisti, 2005).

Les enzymes lipolytiques associées aux détergents sont donc étudiées afin d'en définir et d'en améliorer les performances (Kumar *et al.*, 2009). Dans le cas des lipases, elles doivent fonctionner en présence de détergents et donc répondre à un cahier des charges : une faible sélectivité, un pH optimum compris entre 10 et 11 et une température optimum comprise entre 30°C et 60°C. De plus, dans les lessives, la stabilité des enzymes doit être maintenue en dépit de la présence de protéases et d'agents chimiques capables de provoquer leur dénaturation. Ces paramètres ont permis de fixer des critères pour la recherche de nouvelles enzymes lipolytiques (Bornscheuer and Kazlauskas, 1999). De nombreux produits commerciaux ont été obtenus dont le plus connu est la « Lipolase[™] » commercialisée par la firme Novozymes[®] (Danemark). Cette lipase provient de *Humicola lanuginosa* (aussi appelé *Thermomyces lanuginosus*). Elle est produite dans *Aspergillus oryzae* depuis 1994. L'année suivante, deux nouvelles lipases sont introduites sur le marché, « Lumafast[™] » produite par *Pseudomonas mendocina* et « Lipofast[™] » produite par *Pseudomonas alcaligenes*. Elles sont commercialisées par Genencor International (aujourd'hui propriété de DuPont[®]) (Jaeger and Reetz, 1998).

1.5.3. Applications dans l'industrie textile et papier

Jusqu'à présent, afin de réduire la matière grasse des peaux animales, des produits chimiques étaient utilisés. Pour réduire leur usage et l'impact environnemental, des lipases et d'autres hydrolases sont maintenant privilégiées (Hasan *et al.*, 2006). L'usage de lipases est également observé pour assouplir des tissus et favoriser leur teinture (Fickers *et al.*, 2008).

Lors de la production de papier, des résidus accompagnent généralement les principaux composants du bois. Ces résidus sont en général hydrophobes et sont principalement des triglycérides. Leur élimination demande l'usage de beaucoup de produits chimiques et nécessite du temps. L'utilisation d'enzymes lipolytiques a, dans certains cas, permis d'éliminer 90 % des contaminants (Jaeger *et al.*, 1994). Au Japon, l'entreprise "Nippon Paper Industries" utilise notamment un procédé à partir de la lipase de *Candida rugosa* (Jaeger and Reetz, 1998).

1.5.4. Applications dans l'industrie agroalimentaire et laitière

L'industrie des produits alimentaires a une large utilisation des enzymes lipolytiques, autant dans leur fonction d'hydrolyse que dans leur fonction de synthèse. La position, la longueur de chaîne acylée

tout comme le degré d'insaturation sont autant de facteurs modifiant les propriétés physiques, nutritives ou gustatives d'un triglycéride présent dans un aliment. Les lipases peuvent modifier les propriétés de ces lipides. Par exemple, l'acide palmitique contenu dans le beurre de cacao peut être remplacé par de l'acide stéarique par transestérification, ce qui permet d'augmenter le point de fusion (Sharma *et al.*, 2001).

Les lipases peuvent également réaliser l'interestérification, cette propriété est utilisée lors du processus de fabrication de la margarine. D'autres aliments enrichis permettent l'apport d'acides gras polyinsaturés (AGPI) dont les impacts positifs sur la santé sont bien connus maintenant (anti-inflammatoire et protection contre les maladies cardio-vasculaires). La production d'aliments enrichis peut faire appel à des enzymes lipolytiques afin d'obtenir des lipides enrichis spécifiquement (Yamane *et al.*, 1992). Le choix des positions peut être fait pour que la lipase pancréatique puisse facilement les hydrolyser et donc faciliter leur assimilation et par exemple leur rôle protecteur pour l'organisme consommateur.

L'industrie agroalimentaire utilise également les enzymes lipolytiques pour dégraisser les viandes et les chairs de poisson (Seitz, 1974; Bornscheuer and Kazlauskas, 1999). Les réactions d'hydrolyse peuvent provoquer la libération d'acides gras volatils (chaîne carbonée de moins de six carbones). Ces acides gras volatils permettent l'obtention d'arômes et de saveurs dans de nombreux produits alimentaires comme le vin, le saucisson, ou le thé noir (Pandey *et al.*, 1999).

De nombreux breuvages, soupes, sauces, requièrent au cours de leur fabrication différentes enzymes dont des lipases. L'industrie laitière est également consommatrice d'enzymes lipolytiques pour améliorer les propriétés de ses produits. Il faut, en effet, développer à la fois les saveurs mais également les odeurs des produits laitiers, comme les fromages. Les lipases permettent d'accélérer la maturation des fromages et peuvent également favoriser la dégradation de crèmes et des graisses présents dans le lait (Saxena *et al.*, 1999). Les lipases sont également utilisées dans la production de pains et de gâteaux.

L'usage d'enzymes lipolytiques dans les industries agroalimentaires ne se limite pas à la fabrication des produits de consommation mais peut également diminuer l'impact des unités de production sur l'environnement. Les déchets de ces industries contiennent souvent des lipides et des graisses et cela peut provoquer le colmatage de canalisations. L'addition de lipases et de protéases aux effluents permet de limiter ces phénomènes.

1.5.5. Autres applications

Les lipases peuvent être utilisées en diagnostic médical pour quantifier les triglycérides. Cette méthode, très précise, libère le glycérol par l'action d'une lipase qui est ensuite quantifiée par méthode chimique ou enzymatique. Des lipases peu spécifiques sont en général utilisées pour ces dosages (Pandey *et al.*, 1999).

Des lipases et d'autres enzymes lipolytiques sont également utilisées comme biomarqueurs pour déterminer l'origine de certaines pathologies. Les micro-organismes pathogènes produisent souvent des enzymes lipolytiques spécifiques qu'il suffit de détecter et de quantifier pour déterminer le lien entre un micro-organisme et des symptômes (Hasan *et al.*, 2006). L'identification de la bactérie *Staphylococcus aureus* peut être réalisée sur le milieu Baird-Parker. Ce milieu contient notamment des extraits de lécithine permettant l'action d'une lecithinase propre à *S. aureus* (Matos *et al.*, 1995).

Les enzymes lipolytiques sont par ailleurs utilisées dans l'industrie pharmaceutique, notamment pour leurs propriétés de sélectivité. La production d'énantiomères de molécules très spécifiques est possible par l'usage d'enzymes, là où la production par voie chimique peut provoquer l'obtention de mélanges racémiques inutilisables comme molécules d'intérêt. Certaines molécules utilisées contre le virus du SIDA disposant de propriétés chirales particulières sont produites de cette manière (Jaeger and Eggert, 2002). Dans le domaine cosmétique enfin, certaines fragrances de parfum sont obtenues par l'utilisation d'enzymes lipolytiques (Pandey *et al.*, 1999).

1.5.6. La production des monoglycérides

Divers domaines industriels (alimentaire, cosmétique, pharmaceutique) utilisent les monoglycérides (MG) comme émulsifiants. Les lipases ont ainsi été étudiées pour leur capacité à générer des MG. Malheureusement, la réaction enzymatique catalysée par les lipases conduit souvent à l'obtention d'un mélange de monoglycérides et de diglycérides (Bornscheuer, 1995). Bien que la production de MG par la lipase de *Penicillium roqueforti* possède un rendement de 90% en une seule étape de glycérolyse, le procédé n'est pas viable d'un point de vue économique. Par conséquent, les procédés enzymatiques ont du mal à concurrencer les procédés chimiques utilisés actuellement. La voie enzymatique est cependant préférée pour la synthèse de MG contenant des substituants labiles.

1.5.7. Enzymes de microalgues et applications énergétiques

Actuellement, l'association du terme « microalgue » au monde industriel permet d'explorer quatre domaines que sont l'énergie, l'aquaculture, les cosmétiques et la nutrition. L'utilisation des

microalgues comme source d'enzymes innovantes a été, durant de nombreuses années, sous-exploitée par comparaison aux bactéries et levures. Ce sous-chapitre va se focaliser principalement sur l'usage des microalgues dans le domaine des biocarburants qui est en vaste expansion depuis ces dernières années. La production de biocarburants est basée sur l'utilisation des lipides. Les microalgues représentent par rapport à leurs homologues oléagineux terrestres, un potentiel de production de matières premières de dix à trente fois supérieur par unité de surface (Cadoret and Bernard, 2008). L'intérêt des microalgues dans ce domaine peut notamment être le recyclage du CO₂ et l'utilisation de l'énergie solaire. Lors de nombreuses études, il a été démontré que le stress appliqué à certaines espèces de microalgues provoque le changement de leur composition lipidique. Ce changement de composition est étudié pour en détecter les causes et les effecteurs, notamment d'éventuelles enzymes lipolytiques (Franz *et al.*, 2013).

La composition des lipides est un paramètre important. Les recherches sont actuellement orientées vers l'augmentation des teneurs en lipide et leur enrichissement en acides gras choisis. Cette démarche est déjà exploitée chez la bactérie *Escherichia coli* (Zhang *et al.*, 2011; Zheng *et al.*, 2012). Les techniques actuelles de production de biodiesel passent par une transesterification chimique en milieu alcalin. Néanmoins, la transesterification est également une des réactions catalysées par des lipases et elle a déjà été utilisée dans d'autres procédés industriels. L'opportunité d'associer la composition spécifique en acide gras et des enzymes lipolytiques spécifiques issues de microalgues a donc été envisagée (Radakovits *et al.*, 2010). Sur cette base, la microalgue *Phaeodactylum tricornutum* a, par exemple, été transformée avec l'un de ses propres gènes codant pour une thioestérase et cela a permis d'augmenter la quantité de certains acides gras (Gong *et al.*, 2011).

2. Caractériser une activité lipolytique par biochimie et biologie moléculaire

2.1. Préambule

Il existe de nombreuses méthodes pour mettre en évidence ou caractériser une activité lipolytique. Dans ce chapitre, les méthodes biochimiques sont présentées dans une première partie. Une seconde partie est consacrée à la prédiction de la fonction lipolytique par analyse bioinformatique. La dernière partie s'est focalisée sur l'étude de la microalgue *Isochrysis galbana*.

2.2. L'approche biochimique

Expérimentalement, la recherche d'une fonction enzymatique peut-être réalisée par l'observation de la disparition d'un substrat ou de l'apparition d'un produit. Au paragraphe 1.2.2 les modèles permettant de quantifier l'activité lipolytique ont été décrits. La qualité de l'interface conditionne la conformation de l'enzyme et son activité (Panaiotov and Verger, 2000). Implicitement, la mesure de l'activité est donc affectée (Gargouri *et al.*, 1986). La qualité de l'interface et le contrôle des conditions expérimentales sont importants pour mettre en évidence une activité lipolytique (notamment pour les lipases et phospholipases). Cette recherche dépend aussi de l'environnement physico-chimique de l'enzyme. Celui-ci ne doit pas être inhibiteur de l'activité de l'enzyme.

2.2.1. Méthodes en milieu de culture solide

La mise en évidence d'une activité lipolytique sur gélose est une méthode usuelle pour sélectionner un organisme produisant une telle activité. L'hydrolyse du substrat inclus dans la gélose entraîne un changement d'aspect du milieu. Les premiers substrats utilisés ont été la tributyrine ou la trioléine. L'émulsion du triglycéride crée un aspect opaque ou trouble du milieu de culture gélifié. L'activité lipolytique est démontrée par l'apparition de zones éclaircies résultant de l'obtention de glycérol et d'acide butyrique ou oléique soluble dans l'eau. Ces zones (halos) apparaissent autour des colonies ou autour d'un dépôt d'extraits cellulaires après un certain temps d'incubation. La dimension du halo est proportionnelle à la concentration d'enzyme présente (Thomson *et al.*, 1999; Singh *et al.*, 2006). Il est possible d'utiliser du Tween comme substrat, c'est un composé de polyoxyéthylène–sorbitan lié à un mono–ester dont la longueur varie (monolaurate pour le Tween 20, monooléate pour le Tween 80). Les halos d'hydrolyse, sont en présence de chlorure de calcium, opaque par formation de sels calciques insolubles (Sierra, 1957). La distinction de zones d'hydrolyse peut parfois être difficile pour certains substrats. L'ajout d'indicateurs colorés est ainsi devenu courant pour le Tween, la lécithine de soja et même pour la trioléine. Ces colorants sont nombreux et les plus utilisés sont le bleu Victoria, le bleu Spirit, le bleu Nil ou le bleu Nuit (Shelley *et al.*, 1987). Le changement de pH

provoqué par la libération des acides gras provoque le virage des indicateurs. C'est la technique de base pour détecter une activité lipolytique. L'analyse peut cependant être biaisée car un microorganisme en croissance peut produire des acides issus du métabolisme des sucres (Rahman *et al.*, 2007).

Un autre marqueur, la Rhodamine B, peut être ajouté pour détecter une activité lipolytique sur la trioléine ou le Tween. La détection est alors basée sur la formation de complexes entre la Rhodamine B et les acides gras libérés. Ces complexes sont révélés par fluorescence sous rayonnement UV à 350 nm (Samad *et al.*, 1989; Jette and Ziomek, 1994). Des halos oranges fluorescents sont visualisés (Kouker and Jaeger, 1987). Il existe aussi une proportionnalité entre la dimension du halo et la concentration d'enzyme.

Pour détecter des activités phospholipasiques, Chrisope *et al.* (1976) ont mis au point une méthode utilisant la lécithine couplée au bleu Nil sur des milieux gélosés. Le milieu obtenu a permis de détecter des activités phospholipase C produisant des précipités de diacylglycérols insolubles et des activités phospholipase A produisant des zones d'éclaircissement.

Ces milieux gélosés peuvent être également utilisés pour réaliser un zymogramme. L'originalité provient de la séparation préalable de l'échantillon protéique par migration électrophorétique. L'objectif est de pouvoir déceler une activité lipolytique identifiable pour une bande protéique donnée après séparation électrophorétique. Le transfert de l'enzyme peut être réalisé par capillarité ou par électro-transfert en appliquant la technique de Western blot (Sinnott *et al.*, 1998; Pan *et al.*, 2011).

2.2.2. Méthodes par titration

Les réactions lipolytiques libèrent une quantité d'acide qui peut être mesurée par titration. La méthode communément appelée pH-stat est une des méthodes les plus courantes (Lingane, 1948). L'émulsion du substrat est maintenue par agitation et par l'ajout de gomme arabique. L'ajout de soude titrée pour maintenir un pH constant est mesuré et permet d'évaluer l'activité d'hydrolyse. Cette méthode est suffisamment sensible pour évaluer des activités comprises entre 0,1 et 1 μ mole d'acides gras libérés par min (Jaeger *et al.*, 1994). L'huile d'olive et la tributyrine sont utilisées comme substrat. La caractérisation de l'activité spécifique d'une estérase de la plante *Jatropha curcas* illustre cette technique (Staubmann *et al.*, 1999).

2.2.3. Méthodes colorimétriques, turbidimétrique et fluorimétrique

Les méthodes colorimétriques utilisent essentiellement les esters de p-nitrophényle comme substrats. L'activité de l'enzyme est déterminée par la libération de p-nitrophénol mesurée par colorimétrie à 410 nm (Winkler and Stuckmann, 1979). La variation des longueurs de chaîne des différents esters permet d'affiner la spécificité de l'enzyme. D'autres études utilisent des esters liés au β -naphthyl, l'hydrolyse libère du β -naphthol mesurable par colorimétrie à 560 nm (Nachlas and Blackburn, 1958). Ces substrats permettent de détecter des activités carboxyl ester hydrolase. Dans le cas d'une activité thioestérase, ces substrats montrent des limites. Shahi *et al.* (2006) ont notamment montré que l'utilisation de chaînes carbonées greffées à un coenzyme A constituent une alternative. Leur étude a montré l'hydrolyse de palmitoyl-CoA et de stearoyl-CoA mais aucune activité n'a été détectée pour les esters de p-nitrophényle correspondants (-palmitate ou -stéarate). Yamada *et al.* (1998) ont pour leur part testé l'acide dithio-bis nitrobenzoïque (DTNB) comme substrat. L'activité thioestérase est mise en évidence par la libération d'acide thionitrobenzoïque (TNB) dont l'absorbance est mesurée à 415 nm.

Il est également possible de suivre l'activité enzymatique par turbidimétrie. Les acides gras produits par l'hydrolyse enzymatique du Tween 20 provoquent un précipité en présence de chlorure de calcium dont l'intensité est mesurée par turbidimétrie. Cette méthode s'est révélée plus sensible que la turbidimétrie sur esters de p-nitrophényle ou que le pH-stat avec du Tween 20 lors de la caractérisation d'une carboxylestérase de *Lysobacter enzymogen* (von Tigerstrom and Stelmaschuk, 1989).

Concernant les méthodes fluorimétriques, la Rhodamine B est le principal marqueur employé. L'interaction de la Rhodamine est possible avec de nombreux substrats et elle peut être aisément appliquée à grande échelle (Jette and Ziomek, 1994). D'autres complexes associant du pyrène ont servi pour détecter des activités du type lipase ou phospholipase. Dans ce cas, une des chaînes acylées est remplacée par un résidu pyrène fluorescent. Un analogue de phospholipide est obtenu et utilisable par fluorimétrie. Thuren *et al.* (1988) ont utilisé cette méthode pour la caractérisation d'une phospholipase A2. Une autre étude a utilisé un triglycéride naturellement composé d'un ester fluorescent, l'acide parinarique. Cette méthode a permis la détection des activités lipolytiques d'extraits de graines d'*Arabidopsis thaliana* (Beisson *et al.*, 1999). Les principaux avantages des méthodes fluorimétriques sont leur sensibilité pour la détection d'activités lipolytiques notamment pour des enzymes non purifiées. La fluorescence permet d'une part l'analyse même en présence d'un

trouble apporté par l'échantillon (par exemple dans le cas d'extraits bruts) et d'autre part la fluorescence offre une meilleure stabilité à des valeurs élevées de pH et de température (Wahler *et al.*, 2001; Gilham and Lehner, 2005).

2.2.4. Méthodes chromatographiques

Trois principales méthodes de chromatographie sont utilisées pour caractériser une activité lipolytique, par l'analyse des acides gras libérés : la chromatographie sur couche mince (CCM), la chromatographie en phase liquide à haute performance (HPLC) et enfin la chromatographie en phase gazeuse (CPG).

Parmi ces méthodes, la plus simple à mettre en oeuvre est la CCM. L'étude de *Corynebacterium acnes* a révélé une activité lipolytique par analyse de sébum en présence de la bactérie par comparaison avec un témoin sans bactérie. Une CCM a permis de mettre en évidence la présence d'acides gras libres résultant de l'hydrolyse partielle des triglycérides dans l'échantillon contenant les bactéries (Townsend *et al.*, 1999). Une méthode utilisant des substrats radiomarqués a également permis, par CCM, de déterminer la spécificité de la lipoprotéine-lipase (Jiang and Mannervik, 1999). Plus tard, l'étude d'une lipase du colza (*Brassica napus*) a également utilisé comme substrat la trioléine radiomarquée. L'activité lipolytique de la lipase a été démontrée par la migration et la quantification des différents produits de dégradation par CCM (Hills and Mukherjee, 1988). Récemment, au sein du laboratoire, les premières hypothèses concernant la présence d'activités lipolytiques au sein de la microalgue *Isochrysis galbana* ont été émises. L'action d'un extrait protéique d'*Isochrysis galbana* sur les lipides totaux de cette microalgue a provoqué l'apparition d'acides gras libres et la réduction voire la disparition des spots correspondant à certaines classes de lipide (Devos M, Thèse 2005).

Des méthodes HPLC ont été développées pour évaluer des activités lipolytiques. En 1989, Ergon et André ont mesuré une activité lipasique en analysant les produits issus de l'hydrolyse enzymatique de différentes classes de lipide. Une autre méthode a été mise en oeuvre en utilisant différents esters de β -naphtyl. Ceux-ci, une fois hydrolysés, forment du naphthol qui a été dosé par chromatographie en phase inverse (Maurich *et al.*, 1991).

La séparation et la quantification des produits d'hydrolyse de différentes enzymes lipolytiques peuvent être faites par CPG. La méthodologie consiste à convertir les acides gras en acides gras méthylés pouvant alors être dosés en CPG. Bereuter and Lorbeer (1995) ont analysé les produits d'hydrolyse des différents triglycérides. Ils ont ainsi pu les identifier et les quantifier. Par l'utilisation

de détecteurs plus performants, la CPG se révèle plus sensible que les méthodes HPLC. Cette différence a actuellement tendance à diminuer principalement par l'utilisation du spectromètre de masse associé à la chromatographie (Beisson *et al.*, 2000; Christie and Han, 2010). Mu *et al.* (2000) ont utilisé la CPG afin d'analyser l'activité de la lipase immobilisée de *Rhizomucor miehei*. L'activité catalysée dans ce cas a été une activité d'interestérification, la CPG a permis de déterminer précisément la composition des triglycérides. Récemment, la CPG a également permis d'observer l'enrichissement en AGPI de fractions lipidiques caractérisant ainsi la spécificité d'hydrolyse de différentes lipases (Rupani *et al.*, 2012).

La CPG peut être utilisée pour obtenir le profil d'acides gras de lipides totaux extraits d'un organisme. Elle peut donc caractériser l'activité *in vivo* de certaines enzymes lipolytiques. C'est le cas par exemple, des thioestérases qui peuvent modifier le profil d'acides gras dans les lipides totaux d'organismes exprimant de manière endogène ou hétérologue ce type d'enzyme (Zheng *et al.*, 2012). La CPG a récemment permis de caractériser des thioestérases d'organismes oléagineux (plantes terrestres et microalgues) (Zhang *et al.*, 2011; Chen *et al.*, 2012)

2.2.5. Méthodes immunologiques et radioactives

Les méthodes immunologiques ne sont généralement pas utilisées pour mesurer une activité lipolytique mais plutôt pour quantifier ou établir la présence d'une enzyme lipolytique connue dans un environnement biologique. Cette détection est indépendante de l'activité lipolytique, puisqu'elle peut être réalisée autant sur la forme active qu'inactive de l'enzyme (Beisson *et al.*, 2000). L'obtention d'anticorps spécifiques de l'enzyme est réalisée en amont par une méthodologie classique. Une fois obtenus, les anticorps sont souvent utilisés par technique ELISA et pour la localisation cellulaire (De Caro *et al.*, 1999).

Plusieurs approches utilisent des chaînes d'acides gras radiomarqués. Ces techniques sont spécifiques et très sensibles, mais demeurent complexes à mettre en oeuvre et sont assez peu adaptées à l'analyse de nombreux échantillons (Briquet-Laugier *et al.*, 1999). Toutefois, de par sa sensibilité, cette méthode est particulièrement adaptée et utilisée pour la détection et la caractérisation d'activité enzymatique de thioestérases sur des acyl-CoA radiomarqués (Jha *et al.*, 2010; Liu *et al.*, 2010).

2.2.6. Autres méthodes

D'autres approches existent pour déterminer une activité lipolytique. Certaines méthodes utilisent les particularités des lipases qui sont actives uniquement aux interfaces. La mesure de la pression de

surface est l'un des critères modifiés lors d'activités lipolytiques. La structure de l'interface eau/lipide est en effet réarrangée par l'action de l'enzyme (densité et pression de surface) (Verger, 1980). Cette technique est identifiée sous le nom de technique des films monomoléculaires. En 1995, l'association de la technique ELISA à la technique des films monomoléculaires a permis de mesurer l'activité lipolytique de la lipase gastrique humaine et notamment l'activité spécifique à l'interface eau-lipide (Aoubala *et al.*, 1995).

La technique de la microgoutte d'huile est une autre méthode permettant d'évaluer les propriétés de l'interface pendant une activité lipolytique. Cette méthode permet de visualiser l'effet d'un changement de tensions de surface. En milieux aqueux, en présence d'une activité lipolytique, la forme de la gouttelette est modifiée et visualisée (Labourdenne *et al.*, 1997). L'observation de l'hydrolyse par les enzymes lipolytiques sur une bicouche lipidique peut également s'effectuer par microscopie à force atomique. Le substrat utilisé, du dipalmitoyl de phosphatidylcholine ou de phosphatidylethanolamine est alors lié à du mica (Beisson *et al.*, 2000; Balashev *et al.*, 2001).

L'ensemble de ces techniques permet de détecter une activité lipolytique aussi bien pour une enzyme connue que dans le cas d'une enzyme à caractériser. Actuellement, la méthode de référence est le pH-stat. Cependant, cette méthode n'est pas facilement adaptable lors de tests multiples ou de balayage à haut débit (Mateos-Diaz *et al.*, 2012). Les approches moléculaires et le développement des études de métagénomique augmentent les conditions et le nombre d'échantillons à tester. Le choix d'une méthode est donc focalisé sur la possibilité de tester un grand nombre d'échantillons tout en conservant le même degré de sensibilité. Les méthodes fluorimétriques et colorimétriques couplées aux plaques multi-puits sont actuellement développées en ce sens (Reyes-Duarte *et al.*, 2012).

2.3.L'approche bioinformatique : gènes, domaines, motifs et séquences consensus des enzymes lipolytiques

Le développement des techniques de séquençage et des bases de données a abouti à la création de nombreux programmes d'annotations et de séquençage pour l'ensemble du monde vivant. Les annotations de génomes, ainsi que les séquençages complets du génome de certaines espèces ont servi au développement de banques de données et d'approches bioinformatiques. Les résultats de ces développements ont permis d'optimiser le clonage et la caractérisation de gènes pouvant coder pour une activité lipolytique.

2.3.1. Les bases de données

Pour pouvoir reconnaître la fonction lipolytique au sein d'une protéine, il faut tout d'abord définir les domaines spécifiques de cette fonction. Des bases de données « généralistes » permettent d'aligner les séquences protéiques sans restriction de manière à orienter la démarche de recherche. En général, les premiers alignements de la séquence protéique sont réalisés sur des bases de données, comme celles de la SwissProt/UniprotKB (Bairoch and Apweiler, 2000) ou du NCBI/RefSeq (www.ncbi.nlm.nih.gov). Dans la majorité des cas, ces bases de données disposent d'outils intégrés permettant l'annotation de domaines.

Le motif le plus retrouvé pour les enzymes lipolytiques est le pentapeptide GX SXG contenant la sérine du site catalytique. Ce pentapeptide a permis de classer les premières lipases et carboxylestérases dans des bases spécifiques. Au fur et à mesure, de nouveaux motifs sont apparus définissant de nouvelles familles d'enzymes lipolytiques comme vu au paragraphe 1.3.3. Il est possible d'aligner une séquence contre ces bases de données spécifiques. Les séquences protéiques y sont classées en fonction de leur origine, ou en fonction de leur activité type ou encore de la conformation de certains domaines peptidiques.

Lipase Engineering Database (LED)

La LED est une base de données spécifique des enzymes lipolytiques. Le classement s'effectue en différentes familles sur la base d'homologies de séquences établies à partir d'alignements multiples globaux. La LED constitue un instrument d'analyse des relations existant entre les séquences, les structures et les fonctions des lipases, carboxylestérases et protéines proches (Pleiss *et al.*, 2000).

En 2004, la LED rassemblait 3148 séquences correspondant à 2313 protéines dont 35 % de protéines hypothétiques et moins de 5 % de protéines structurellement caractérisées (Fisher, 2004). Aujourd'hui, la LED dénombre 24783 séquences pour 18585 protéines soit presque 75 % de protéines hypothétiques et environ 4,5% de protéines structurellement caractérisées. Les séquences sont hiérarchisées en trois niveaux : la séquence elle-même, la famille d'homologues et la superfamille. Les familles d'homologues présentent, entre elles, un faible mais significatif degré de similitude, elles ont été regroupées en 38 superfamilles. Toutes les séquences montrant un haut degré de similitude ont été classées en 112 familles d'homologues. Les superfamilles ne présentent pas de similarités de séquences significatives entre elles (Fischer, 2003). Les protéines ont également été réparties en trois classes, GX, GGGX et Y, selon la séquence et la structure du puits oxyanionique (Pleiss *et al.*, 2000) (Tableau 3).

Classe	Superfamilles	Familles	Séquences	Protéines	Structures tridimensionnelles	Particularités	Exemples
GX	27	82	14905	11402	610	Le résidu X à l'origine du puits oxyanionique est conservé et toujours précédé d'une glycine G.	Lipases de bactéries et champignon, lipase hépatique, pancréatique, gastrique, lipoprotéine-lipase, lipases acides lysosomales.
GGGX	6	22	7325	5278	282	Le résidu X est un acide aminé hydrophobe placé à la suite d'un motif bien conservé GGG.	Carboxylestérases, lipases activées par des sels biliaires, lipases hormono-dépendantes, acétylcholinestérases et des thioestérases.
Y	5	8	2553	1905	225	Le puits oxyanionique est formé par le groupe hydroxyle de la tyrosine, résidu hautement conservé.	Lipases de <i>Candida</i> sp.

Tableau 3: Données générales de la banque de données LED

Microbial Esterase et Lipase Database (MELDB)

La MELDB est une base de données spécifique des lipases et des carboxylestérases microbiennes. Cette classification utilise les relations entre les propriétés biochimiques et les régions fonctionnelles conservées. Dans cette étude, les enzymes ont été classées d'après des alignements locaux sur les régions conservées. Ces régions ont été divisées en deux groupes, les domaines connus et les régions conservées de fonction inconnue. Cette approche diffère de celles qui, traditionnellement, se fondent sur des alignements globaux et des calculs de distance. Elle permet de réduire les méprises dues aux mauvais alignements de régions éloignées ou différentes. La MELDB recense 883 carboxylestérases et lipases, réparties respectivement en 20 et 30 familles. Les groupes générés par la MELDB correspondent globalement à ceux de la LED. Toutefois, des différences significatives ont été observées dans la composition des sous-groupes. Les lipases issues de différentes espèces de *Bacillus* sp. se retrouvent ainsi séparées dans deux sous-groupes de la MELDB. De la même manière, des lipases issues de *Moraxella* sp. se retrouvent en deux sous-groupes dont l'un est composé de lipases issues de bactéries pathogènes (Kang *et al.*, 2006).

ESTerases and alpha/beta Hydrolase Enzymes and Relatives (ESTHER)

Cette base de données, créée en 1996, est dédiée aux enzymes avec un mode de repliement de type hydrolase α/β (Cousin *et al.*, 1996). Elle est composée de 4 groupes C, H, L et X pour 34504 séquences. Ces 4 groupes représentent 86 super-familles composées de 155 familles d'enzymes. L'organisation de ces groupes est dépendante des similarités de séquence. Elle est basée sur des informations biologiques et s'est constituée par rapport aux familles de protéines issues de bases de données externes (Pfam, Interpro, Uniprot). Cette base de données regroupe le maximum de données concernant les séquences protéiques d'ester hydrolases ayant une cystéine ou une sérine comme résidu nucléophile dans le site catalytique. Le groupe C représentant 17 % de cette base de données intègre notamment majoritairement les carboxylestérases. Le groupe H inclue les lipases hormono-sensibles et le groupe L inclue la majorité des autres lipases. Ces deux groupes représentent 7,5 % de l'ensemble de la base de données. Enfin, le groupe X est majoritaire avec 75,5 %, il intègre notamment les cutinases et thioestérases mais également de nombreuses hydrolases α/β qui non caractérisées (Lenfant *et al.*, 2013).

Thioester-active Enzymes (ThYme)

Cette base de données est dédiée aux enzymes impliquées dans la synthèse d'acides gras et de polyketides, notamment à partir de substrats contenant une liaison thioester. Les polyketides représentent un groupe de composés très diversifiés dont l'architecture de base est une chaîne acylée (Chan *et al.*, 2009). La base de données ThYme est composée de 23 familles dont 8 seulement sont des hydrolases α/β . Dans ces familles, la principale fonction retrouvée est Acyl-ACP hydrolase. Cette base de données a été en partie présentée au paragraphe 1.2.3.

Autres bases de données

Afin de compléter les sources utilisées lors d'un travail d'alignement, d'autres ressources sont utilisées. Il s'agit de la Protein Data Bank (PDB) pour les structures de protéines (<http://www.rcsb.org/>) et de Pfam une banque de données constituée de familles de protéines obtenues par alignement de séquences (<http://pfam.sanger.ac.uk/>). Une dernière base de données, Prosite, est désignée pour la recherche de fonctions. Les domaines et motifs conservés au sein des protéines y sont définis (<http://prosite.expasy.org/>). Le tableau 4 présente quelques exemples de motifs correspondant à des enzymes lipolytiques.

Motifs	Description	Type Enzymatique	Numéro d'accès
F-[GR]-G-X(4)-[LIVM]-X-[LIV]-X- G-X-S-[STAG]-G <i>S est la sérine active dans le site catalytique</i>	Acides aminés retrouvés en amont du site catalytique chez les carboxylestérases	Carboxylestérase	PS00122
[EDA]-[DG]-C-L-[YTF]-[LIVT]-[DNS]-[LIV]-[LIVFYW]-X-[PQR] <i>C est une cystéine impliquée dans la formation d'un pont disulfure</i>	Acides aminés retrouvés en amont du site catalytique chez les carboxylestérases	Carboxylestérase	PS00941
[LIV]-[KG]-[LIVFY]-[LIVMST]- G-[HYWV]-S-[YAG]-G-[GSTAC] <i>S est la sérine active dans le site catalytique</i>	Acides aminés retrouvés dans les lipases à sérine	Lipase	PS00120
[LIVMFYAG](4)-G-D-S-[LIVM]-X(1,2)-[TAG]-G <i>S est la sérine active dans le site catalytique</i>	Acides aminés retrouvés dans les lipases de type "G-D-S-L"	Lipase	PS01098
[LIVMF](2)-X-[LIVMF]-H-G(2)-[SAG]-[FYW]-X(3)-[STDN]-X(1,2)-[STYA]-[HAGFT] <i>La première histidine peut appartenir au site catalytique</i>	Acides aminés retrouvés dans les enzymes lipolytiques de motif "G-D-X-G" comprenant une histidine	Lipase	PS01173
[LIVM]-X-[LIVMF]-[SA]-G-D-S-[CAS]-G-[GA]-X-[LI]-[CAVT] <i>S est la sérine pouvant appartenir au site catalytique</i>	Acides aminés retrouvés dans les enzymes lipolytiques de motif "G-D-X-G" comprenant une sérine	Lipase	PS01174
[LIVMA]-C-[LIVMFYWPCT]-C-D-[GS]-[G]-[N]-X-[QS]-C <i>D est l'acide aspartique dans le site catalytique</i>	Acides aminés retrouvés dans le site actif de Phospholipases A2	Phospholipase	PS00119
C-C-[P]-X-H-[LGY]-X-C <i>H est l'histidine dans le site catalytique</i>	Acides aminés retrouvés dans le site actif de Phospholipases A2	Phospholipase	PS00118
H-X-K-X(4)-D-X(6)-G-S-X-N	Acides aminés correspondant au motif 'HKD' retrouvés dans les Phospholipases D	Phospholipase	PS50035
P-X-[STA]-X-[LIV]-[IVT]-X-[GS]-G-Y-S-[QL]-G <i>S est la sérine active dans le site catalytique</i>	Acides aminés retrouvés dans le site actif de la Cutinase 1 à sérine	Cutinase	PS00155
C-X(3)-D-X-[IV]-C-X-G-[GST]-X(2)-[LIVM]-X(2,3)-H <i>D et H sont des résidus du site actif</i>	Acides aminés retrouvés dans le site actif de la Cutinase 2	Cutinase	PS00931

Tableau 4: Quelques motifs retrouvés dans les enzymes lipolytiques à travers la base de données ProSite

Le motif GX SXG classiquement retrouvé chez les enzymes lipolytiques est illustré dans les numéros d'accès PS00122 et PS00120 (en gris). Il existe une variabilité pour la composition non seulement du pentapeptide mais également des séquences bordantes du site actif. Le motif AX SXG est retrouvé dans les lipases de *Bacillus subtilis* (Soliman *et al.*, 2007), ce motif devient TX SXG dans le cas de la lipase de *Candida antarctica* (Pleiss *et al.*, 2000). Enfin pour certaines carboxylestérases, le motif peut être GX SQG voire GS SQG (Vincent *et al.*, 2003; Souleyre *et al.*, 2011).

2.3.2. Identification d'enzymes lipolytiques par homologie de séquences

Les séquences à identifier sont généralement issues de séquences génomiques ou de séquences exprimées (EST). La fonction à associer à une nouvelle séquence est identifiée en comparant la séquence à l'ensemble des séquences présentes dans une base de données. Si une fonction précise n'est pas identifiée, l'appartenance à une classe d'enzymes peut être recherchée par comparaison à des motifs protéiques de fonctions connues. Dans le cas des enzymes lipolytiques, la grande variation des critères d'identification peut rendre particulièrement difficile voire erronée cette identification. De plus, certaines enzymes lipolytiques peuvent avoir été identifiées dans plusieurs familles ou classes enzymatiques (Alvarez-Macarie *et al.*, 1999).

La première méthodologie consistant à comparer la séquence candidate à celles présentes dans des bases de données est réalisée par le biais d'un alignement local ou BLAST (Basic Local Alignment Search Tool) (Altschul *et al.*, 1990). Ce logiciel en ligne est accessible à l'adresse (<http://www.ncbi.nlm.nih.gov/BLAST/>). Une valeur E est calculée permettant d'évaluer la probabilité que l'alignement soit le fruit du hasard et son éventualité biologique. Plus la valeur E est proche de zéro, plus le résultat de la comparaison est pertinent. Le niveau d'analyse peut également se situer du point de vue de la structure secondaire ou de la structure tertiaire.

La seconde méthodologie consiste à rechercher les motifs protéiques d'une fonction donnée (base de données ProSite). La prédiction se réalise sur la base de comparaisons et de maximum de proximité entre deux séquences. Ce sont ces enchaînements de motifs protéiques ou leur emplacement dans la structure qui ont permis une classification récente des thioestérases (Cantu *et al.*, 2010).

2.4. Identification et caractérisation d'une fonction lipolytique chez *Isochrysis galbana*

2.4.1. Indices biochimiques

Actuellement, la recherche d'une activité enzymatique d'intérêt se fait selon deux principales stratégies. Soit elle est caractérisée directement depuis l'organisme modèle, soit il faut cloner le gène et produire l'enzyme dans un autre organisme. Lorsque la culture est réalisable, il est également possible d'obtenir de premières informations biochimiques sur l'activité enzymatique avant un éventuel clonage. Dans le cas de notre organisme modèle, la microalgue *Isochrysis galbana*, une précédente étude au laboratoire consacrée à la composition des lipides totaux a révélé la présence d'acides gras libres (Poisson L, Thèse 1999). Des activités lipolytiques endogènes ont donc été supposées. Lors d'une autre étude, la présence d'activités intracellulaires plus précises, glycolipasiques et phospholipasiques a été supposée (Devos M, Thèse 2005). Enfin, c'est en réalisant l'interaction entre un extrait protéique et un extrait lipidique total issus d'*I. galbana* que l'analyse par CCM a confirmé le travail de 2005. La figure 14 permet de distinguer au moins deux activités. Tout d'abord, les phospholipides et phosphatidylcholines sont partiellement hydrolysés montrant une éventuelle activité phospholipase. Les glycolipides sont également hydrolysés, notamment les digalactosyldiglycérides laissant alors supposer une activité galactolipasique. Une approche moléculaire et bioinformatique a été entreprise pour identifier les gènes pouvant coder pour ces activités lipolytiques chez *Isochrysis galbana*. Cette démarche est venue compléter les résultats obtenus par la méthode biochimique (Godet *et al.*, 2010).

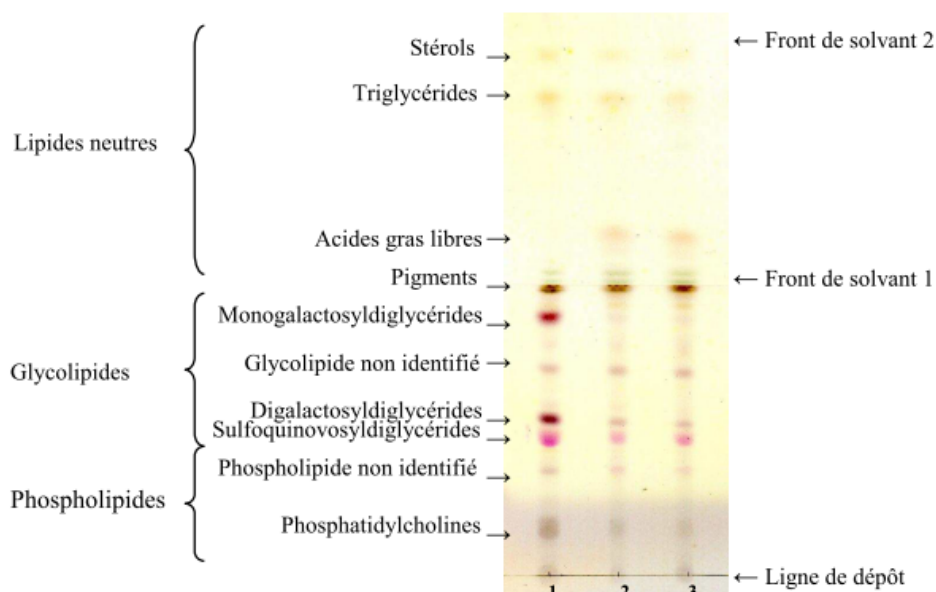


Figure 14: Chromatographie sur couche mince des lipides totaux d'*Isochrysis galbana* pour la mise en évidence de l'activité lipolytique endogène (Godet S, Thèse 2008).

Dépôt 1 : Echantillon témoin, ébouillanté avant incubation à 40 °C pendant 24 heures. Dépôts 2 et 3 : Echantillons non ébouillantés, incubés à 40 °C pendant 24 heures d'agitation.

2.4.2. Les gènes d'*Isochrysis galbana* appartenant au métabolisme lipidique

Isochrysis galbana est une microalgue riche en AGPI notamment l'acide docosahexaénoïque (DHA). Les études de caractérisation de gènes se sont principalement portées sur les enzymes impliquées dans la synthèse de ces AGPI d'intérêt. La biosynthèse de ces acides gras est assurée par différentes enzymes appartenant au métabolisme lipidique, notamment les élongases et les désaturases comme le montre la figure 15. Qi *et al.* (2002), ont produit par expression hétérologue une élongase *IgASE1* fonctionnelle dans la levure. En absence de données génomiques, l'obtention de l'ADN complémentaire à l'origine du clonage a été réalisée avec des amorces dégénérées ciblant le site catalytique de l'élongase. L'expression de cette enzyme a permis l'obtention d'acide linoléique (C18:2n-6, $\Delta 9,12$) et de l'acide α -linolénique (C18:3n-3, $\Delta 9,12,15$) dans la levure. Dans la figure 15 cette enzyme est notée « $\Delta 9$ elo ». Dans une autre étude, Pereira *et al.* (2004) ont isolé le gène codant une désaturase la $\Delta 4$ -désaturase. Ils ont réalisé une librairie d'ADN complémentaires qu'ils ont séquencé. Ils ont détecté la présence d'un motif consensus appelé « His-box » identifiant la désaturase annotée *IgD4* puis *IgFAD4-1*. L'enzyme est notée « $\Delta 4$ des » dans la figure 15.

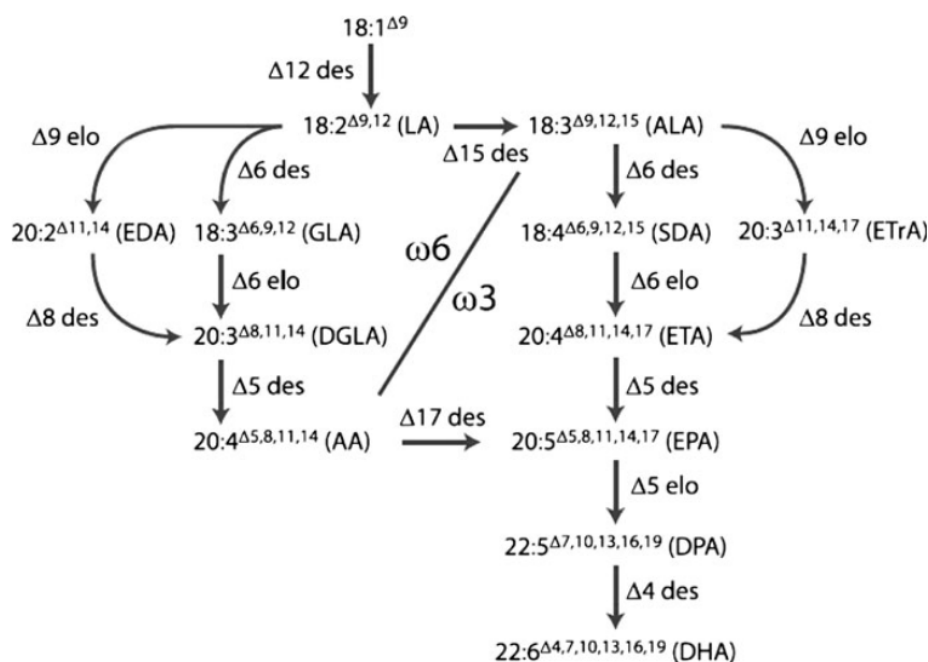


Figure 15: Voie de synthèse supposée pour les AGPIs dans les microalgues (Shi *et al.*, 2012).

« elo » est un acronyme pour élongase, « des » est celui pour désaturase.

Plus récemment, trois enzymes ont été caractérisées par clonage depuis *I. galbana*. Tout d'abord, une nouvelle élongase a été exprimée dans la levure *S. cerevisiae*. La séquence codante de cette enzyme identifiée *IgASE2* montre 83 % de similarité avec la séquence codante *d'IgASE1*. La co-expression avec une désaturase du protiste *Euglena gracilis* a permis de reconstituer une des voix de synthèse des acides gras polyinsaturés (Li *et al.*, 2011).

Le second gène isolé est celui d'une désaturase. Sa séquence codante semble correspondre à une désaturase, la séquence protéique correspondante présente plus de 78 % d'identité avec la désaturase préalablement décrite par Pereira *et al.* (2004). Cette nouvelle désaturase, appelée *IgFAD4-2* semble présenter des caractéristiques nettement plus intéressantes avec un taux de conversion depuis l'acide docosapentaénoïque (DPA) (C22:5 Δ7, 10, 13, 16, 19) vers l'EPA supérieur de 37 % à celui observé avec *IgFAD4-1*. L'expression et la caractérisation ont été réalisées chez *S. cerevisiae* (Shi *et al.*, 2012). Enfin, une étude chinoise a permis de caractériser une autre désaturase d'*Isochrysis galbana* intervenant dans la voie de conversion de l'acide dihomogammalinoléique (DGLA) (C20:3 Δ8, 11, 14) en acide arachidonique (AA) (C20:4 Δ5, 8, 11, 14). Cette enzyme a été définie et notée *IgD5* (Sun *et al.*, 2012).

Concernant les enzymes lipolytiques d'*I. galbana*, deux études ont été réalisées au laboratoire. Une base de données de séquences exprimées annotées (EST) établie pour cette microalgue a été utilisée

(O'Brien *et al.*, 2007). Cette base de données contient un ensemble de transcrits incomplets. L'intégrité des transcrits et la composition des gènes originaux restent à caractériser. Les signaux de début et de fin de traduction sont absents la plupart du temps. Cette base de données contient une préannotation permettant de sélectionner les séquences par rapport à la fonction enzymatique supposée. Pour l'activité lipolytique, le travail développé au laboratoire a permis d'identifier dix séquences codantes dont les clusters correspondants sont présentés dans le tableau 5. Cinq gènes, pouvant correspondre à des gènes de lipases, de carboxylestérases ou de phospholipases, ont été sélectionnés et sont indexés dans les clusters ISL00001076, ISL00001844, ISL00006895, ISL00007519 et ISL00008005 en grisé dans le tableau 5.

Numéro de cluster	Nombre d'EST par cluster	Taille (pb)	Codon stop	Codon start + codon stop en amont	Isolé et cloné au laboratoire	Annotations
ISL00000940	2	655	Non	Non	Non	Esterase/lipase/thioesterase-like
ISL00001076	3	679	Oui	Non	Oui	Phospholipase/Carboxylesterase
ISL00001844	1	533	Non	Non	Oui	Lipase putative
ISL00002069	1	319	Oui	Non	Non	Homologue lysophospholipase
ISL00002977	1	440	Non	Non	Non	Esterase/lipase/thioesterase-like
ISL00003195	2	484	Non	Non	Non	Homologue phospholipase C
ISL00003782	1	413	Non	Non	Non	Lysophospholipase
ISL00006895	1	350	Non	Non	Oui	Lysophospholipase
ISL00007519	1	732	Non	Non	Oui	Carboxylesterase
ISL00008005	1	577	Non	Non	Oui	ACP-Thioesterase

Tableau 5: Annotations des séquences présentes dans la base de données d'EST d'*I. galbana*.

En 2010, par des techniques de RACE/nested-PCR, les extrémités d'une séquence codante indexée au cluster ISL00008005 ont été déterminées. La séquence protéique correspondante a été déduite. L'analyse de séquence a révélé de nombreuses similarités avec des carboxylestérases identifiées notamment chez l'homme et la souris. Le gène candidat a alors été identifié *IgEST1*. La position de la sérine pouvant être impliquée dans le site actif correspond à celle identifiée dans d'autres carboxylestérases. La séquence protéique comporte également de la position 166 à 170 une séquence GX SXG qui est retrouvée dans la majorité des hydrolases à sérine. Parmi les enzymes qui se retrouvent alignées avec *IgEst1*, il est possible de trouver la lysophospholipase I de la souris *Mus musculus*, une carboxylesterase de la plante *Arabidopsis thaliana* ou bien encore une phospholipase de la diatomée *Thalassiosira pseudonana*. La modélisation de ce gène a montré des similitudes avec une acyl protéine thioestérase humaine, permettant notamment d'identifier par simulation de structure

les acides aminés pouvant constituer la triade catalytique (Godet *et al.*, 2010). En utilisant la même démarche, il a été possible de caractériser un autre gène, *IgEst2*, indexé par le cluster ISL00001076.

En 2012, une autre séquence codante indexée dans la base de donnée par le cluster ISL00001844 a été identifiée. L'analyse de séquence a montré l'homologie de séquence avec des lipases et notamment une séquence de lipase non caractérisée provenant d'*Emiliana huxleyi* (39% d'identité) et une séquence de lipase du protozoaire *Trypanosoma cruzi* (18% d'identité). L'analyse de la séquence a permis l'identification des codons initiateur et terminateur et l'obtention de la séquence codante *IgLip*. Cette séquence a également été analysée pour détecter des motifs conservés. Une séquence en particulier a permis de bien identifier des motifs conservés dans *IgLip*, c'est la séquence d'une lipase de *Rhizomucor miehei* (Derewenda *et al.*, 1992). L'analyse a permis l'identification des acides aminés impliqués dans la constitution du site catalytique et confirmé le positionnement du pentapeptide GX SXG entre les positions 196 à 200. De plus, la comparaison avec la lipase de *R. miehei* a permis d'observer un volet amphiphile partiellement conservé (Godet *et al.*, 2012).

Plus récemment, deux gènes correspondants aux clusters ISL00006895 et ISL00007519 ont également été caractérisés. Ces deux gènes ont été analysés et les résultats des études bio-informatiques ont révélé de nombreuses similitudes avec différentes classes d'enzymes lipolytiques connues.

3. Systèmes hétérologues et expression d'enzymes lipolytiques recombinantes

3.1. Préambule

Ce sous-chapitre présente l'expression hétérologue de protéines recombinantes dans les organismes procaryotes (*Escherichia coli*, *Bacillus subtilis*), les levures (*Saccharomyces cerevisiae*, *Pichia pastoris*, *Yarrowia lipolytica*), les cellules d'insectes et les cellules de mammifères. Il existe d'autres organismes de production dont la connaissance et l'utilisation sont moins développées et se limitent parfois à un ou quelques laboratoires. Après une description des principaux hôtes, les différents vecteurs d'expression sont présentés. Enfin, la dernière partie est consacrée à la production d'enzymes lipolytiques.

3.2. La production hétérologue de protéines

3.2.1. Hôte procaryote : *Escherichia coli*

Généralités

Le génie génétique a permis à partir des années 70 la programmation d'une cellule pour produire des protéines d'intérêt. Les premières expressions hétérologues ont été réalisées dans la bactérie à Gram négative *E. coli*. L'expression de la somatostatine humaine reste un exemple classique dans l'utilisation de cette bactérie pour l'expression hétérologue (Itakura *et al.*, 1977). Au début des années 80, l'insuline humaine a été à son tour produite (Riggs, 1981). Cette bactérie présente plusieurs avantages. Elle possède un faible temps de génération (environ 25 min) et le milieu de culture est de composition simple, peu coûteux sans antibiotique. *E. coli* est néanmoins passé de la position d'outil incontournable pour l'expression de protéines, à un outil montrant certaines limites notamment pour les modifications post-traductionnelles.

*Méthodes de production avec *Escherichia coli**

Trois méthodes majeures sont utilisées pour la production de protéines recombinantes chez *E. coli* :

- Expression en protéine de fusion

Le gène d'intérêt est fusionné avec une séquence codant une « étiquette » protéique. Celle-ci permettra l'isolement, l'identification, la visualisation ou encore la solubilisation de la protéine recombinante. L'étiquette protéique peut être, par exemple, une protéine fluorescente permettant la visualisation ou bien un peptide spécifique facilitant la purification ou la solubilisation de la protéine recombinante (Snapp, 2005; Esposito and Chatterjee, 2006).

- Expression périplasmique

La protéine peut être sécrétée dans l'espace périplasmique de la bactérie grâce à la présence d'un peptide signal. L'intérêt de cette technique est de limiter l'impact de l'environnement intracellulaire sur la protéine produite, comme par exemple la dégradation par des protéases intracellulaires. L'expression périplasmique permet également de simplifier les méthodes d'isolement de la protéine produite car celle-ci ne se retrouve pas en interaction avec les résidus de lyses cellulaires. Le risque de contamination par les protéines endogènes ou de dégradation par les processus de cassage mécanique est également réduit. Dans certains cas, la formation des ponts disulfures est favorisée car l'espace périplasmique est un environnement oxydant (Schlegel *et al.*, 2013). Toutefois si l'expression dans le périplasme est régulièrement rapportée elle n'est pas simple. En effet, la protéine exprimée possédant un peptide signal peut se révéler toxique pour la bactérie. L'expression à l'extérieur de la bactérie n'est pas possible chez *E. coli* (Terpe, 2006; Azaman *et al.*, 2010).

- Expression en corps d'inclusion

La surproduction de protéines hétérologues peut entraîner la formation de corps d'inclusion, un des biais de la production de protéines dans *E. coli*. Ce phénomène est multifactoriel et varie d'une protéine produite à une autre. Le mécanisme n'est pas bien connu mais il semblerait que la production de la protéine recombinante soit trop rapide (à cause du temps de génération d'*E. coli*) empêchant ainsi la protéine de prendre une conformation identique à sa forme native. Une protéine mal ou peu repliée laisse des sites d'interactions inter-protéiques accessibles. Ces sites peuvent amener à une aggrégation des protéines entre elles. La formation de ces corps d'inclusion protège néanmoins des enzymes de dégradation telles que les protéases (Tailor *et al.*, 2004). Si ce mode de production est privilégié, il faut après purification resolubiliser la protéine et la rendre active par différents traitements biochimiques (Akbari *et al.*, 2010). Cette méthode chronophage est donc plus coûteuse. Dans certains cas, cette méthode est devenue une stratégie de production de protéines ou d'enzymes d'intérêt (García-Fruitós *et al.*, 2012).

Ces trois méthodes sont retrouvées inégalement dans la littérature. La production cytoplasmique sous forme ou non de protéine de fusion est en effet plus largement utilisée car elle permet d'obtenir une quantité plus importante de protéine recombinante (Terpe, 2006).

Effets de la température et molécules chaperonnes

La production d'une protéine de manière hétérologue génère un stress au niveau de la bactérie. Ce stress provoque différents phénomènes menant dans la majorité des cas à la production de « protéines de réaction au choc », plus communément appelées « Heat-Schock Protein » (HSP) (Ortelli *et al.*, 2003). Ce stress peut se traduire par la production de protéases dégradant la protéine produite. Ce phénomène peut également déstabiliser les plasmides porteurs du gène à exprimer. Le cycle de croissance cellulaire peut être altéré menant parfois à la lyse cellulaire. Globalement, la bactérie met en place, lors de la production hétérologue, une accélération des processus moléculaires et une modification du métabolisme. Ce sont des mécanismes de défense observés dans la bactérie lors de changements d'environnement (Gill *et al.*, 2000). Afin de diminuer ces effets, il est possible de moduler la température du milieu de culture au moment de l'induction et pendant la période de production. La température optimum pour la croissance d'*E. coli* étant de 37 °C, l'abaissement de cette température permet de ralentir le métabolisme cellulaire de la bactérie. La plupart des températures d'expression de protéines dans *E. coli* sont comprises entre 15 et 30 °C (Lawyer *et al.*, 1989). Ce ralentissement favorise une diminution du taux de transcription et de la même manière du taux de traduction permettant à la protéine un meilleur repliement sous forme native (Ruschak *et al.*, 2013). Le changement de température de la culture impose une adaptation du temps de culture (Tableau 6). Récemment, Song *et al.* (2012) ont testé des températures comprises entre 5 à 10 °C et ont ainsi obtenu des protéines recombinantes en diminuant nettement la proportion de corps d'inclusion produits au cours de l'expression.

Température d'incubation (°C)	Temps d'incubation (heures)
8	24-72
15	16-24
20	12-16
25	6-12
30	5-6
37	3-4

Tableau 6: Adaptation du temps d'incubation en fonction de la température de culture (Terpe, 2006)

L'utilisation de molécules chaperonnes est également un moyen d'optimiser la production. En effet, les protéines chaperonnes permettent de faciliter le repliement de la chaîne polypeptidique et d'éviter la formation de corps d'inclusion (Young *et al.*, 2004). Les protéines chaperonnes sont souvent à coexprimer avec la protéine d'intérêt, ce qui peut engendrer des difficultés. D'une part, lorsque la coexpression est réalisée à partir d'un seul plasmide, le ratio entre les deux protéines est alors

difficilement modulable. D'autre part, si la coexpression est réalisée à partir de vecteurs différents, leur maintien dans la cellule hôte impose des contraintes telles que le temps de culture ou la composition du milieu de culture (Makrides, 1996). L'efficacité des protéines chaperonnes semble dépendante de la protéine d'intérêt. La méthode la plus efficace consiste donc en l'obtention à la fois de la protéine et de sa chaperonne issue d'un même organisme (Stephens *et al.*, 2011). Dans certains cas plusieurs protéines chaperonnes sont nécessaires. Elles fonctionnent souvent en réseau de protéines. Ce modèle est néanmoins difficile à reproduire par expression hétérologue. La coexpression de chaperonnes pour chaque nouveau gène candidat est donc à tester.

Enfin, la fonction de la protéine chaperonne peut évoluer en fonction de la température. Spiess *et al.* (1999) ont montré, qu'à faible température, une amylase d'*E. coli* est produite en présence de la protéine chaperonne DegP qui favorise son repliement. Pour des températures plus élevées, l'amylase adopte une conformation inactive et la protéine chaperonne exerce alors une activité protéolytique sur l'amylase. La protéine chaperonne modifie donc son activité en fonction de la température.

Utilisation de protéines de fusion et problématique des ponts disulfures

Le développement de plasmides permettant d'ajouter des étiquettes aux protéines produites a été une évolution majeure pour faciliter l'obtention de protéines recombinantes sous forme soluble. Certaines étiquettes permettent également la détection et/ou la purification de la protéine. L'ajout d'étiquettes peut être réalisé autant à l'extrémité C-terminal qu'à l'extrémité N-terminal de la protéine, voire aux deux (Arnau *et al.*, 2006). Les étiquettes Glutathion S Transférase (GST), thioredoxin (*trx*), Maltose Binding Protein (MBP) et Small Ubiquitin MOdifier (SUMO) permettent une meilleure solubilité des protéines recombinantes. Les étiquettes GST et MBP sont les plus couramment utilisées car elles ont l'avantage à la fois d'augmenter la solubilité mais également de permettre la purification par affinité. Cependant, l'ajout d'une étiquette n'est pas applicable à toutes les protéines. Cette méthode peut en effet, rendre indisponible le site actif ou empêcher le bon repliement de la protéine, enfin elle peut également modifier le point isoélectrique. Freydank *et al.* (2008) ont montré que la présence d'une étiquette de six résidus histidine en C-terminal a rendu le site actif de l'enzyme inaccessible. L'activité a été par contre retrouvée quand l'étiquette a été placée en région N-terminal. Il est possible, dans certains cas de cliver l'étiquette une fois la protéine produite. Cette étape est utile dans le cas de tests d'activité *in vitro* pour observer par exemple l'influence de l'étiquette. Pryor and Leiting (1997) ont montré que la vitesse de la réaction de clivage

Les ponts disulfures, présents à la fois dans les protéines eucaryotes et dans les protéines procaryotes, sont retrouvés en plus grande quantité dans les protéines extracellulaires. La formation des ponts disulfure au sein d'*E. coli* s'effectue majoritairement dans l'espace périplasmique (Andersen *et al.*, 1997). En effet, dans l'espace cytoplasmique est réducteur empêchant la formation de ponts disulfures (Stewart *et al.*, 1998). L'expression hétérologue dans l'espace périplasmique est possible mais le désavantage majeur est la faible quantité de protéines produites. De plus, il est nécessaire d'ajouter un signal peptide pour orienter la protéine dans le périplasma (Gasser *et al.*, 2008). Il est devenu possible d'établir des ponts disulfures directement dans le cytoplasme en utilisant des bactéries génétiquement modifiées. Les gènes concernés sont la thioredoxin réductase (trxB) ou la glutathionne oxido-reductase (gor) responsables de l'environnement réducteur dans le cytoplasme. Les principales souches modifiées sont la souche *E. coli* AD494 dont le gène trxB a été inactivé et les souches Origami® de la société Novagen™ dont les deux gènes trxB et gor ont été inactivés (Bessette *et al.*, 1999; Terpe, 2006).

Une partie des difficultés de l'expression hétérologue dans la bactérie est liée au code génétique. En effet, l'usage des codons n'est pas identique d'un organisme à un autre, et notamment entre procaryotes et eucaryotes (Makrides, 1996). Ces différences se traduisent par des problèmes de mutations et d'arrêt du processus de traduction. Les protéines produites sont tronquées et/ou non fonctionnelles. Le code génétique standard est présenté dans le tableau 7, le code protéique à une lettre designant chaque acide aminé est indiqué en première ligne et les codons initiateurs en seconde ligne.

[illegible]

Au moment de la traduction chaque codon doit être associé à un ARNt recruté pour la synthèse protéique. Le code génétique permet à la cellule d'associer à chacun des 20 acides aminés pouvant constituer les protéines un triplet de trois bases appelées codon. Il existe 64 combinaisons de codons,

61 codent pour un acide aminé. Les trois derniers codons TAA, TGA et TAG ne correspondent à aucun acide aminé et indiquent la fin d'une séquence codante. Le codon méthionine ATG est toujours utilisé pour indiquer le début d'une séquence codante avec alternativement CTG et TTG pour les bactéries (Peabody, 1989). Chaque acide aminé est fixé sur un ARNt qui, grâce à son anticodon, va permettre le décodage du codon correspondant au sein du ribosome. La distribution des différents ARNt est variable d'une espèce à une autre. L'expression hétérologue peut alors se retrouver confrontée à une limitation du nombre d'ARNt pour un codon donné au moment de la traduction. Les codons les plus rares chez *E. coli* sont décrits dans le tableau 8. Les codons arginine et sérine sont particulièrement rares puisqu'ils représentent 2/3 des codons possibles. Pour la proline cette proportion passe à 3/4.

Acides aminés	Codons rares
Arginine	AGG, AGA, CGG, CGA
Leucine	CUA, CUC
Isoleucine	AUA
Sérine	UCG, UCA, AGU, UCC
Glycine	GGA, GGG
Proline	CCC, CCU, CCA
Thréonine	ACA

Tableau 8: Codons les plus rares chez *E. coli*

Pour la production de protéines ou de peptides issus d'organismes eucaryotes, il faut tenir compte de ce phénomène. Lors de la production de la somatostatine en 1977, le gène a ainsi été optimisé en intégrant des codons spécifiques à la bactérie dans sa séquence nucléotidique. Toutefois lors de ces modifications, il faut éviter de créer des zones d'appariement entre bases ou des signaux de terminaison (zones riches en G-C suivies de zones riches en A-T) (Itakura *et al.*, 1977). La production dans la bactérie impose d'évaluer au préalable l'abondance de codons rares dans la séquence d'intérêt. Le problème de rareté des codons semble être plus contraignant pour les codons rares présents à l'extrémité N-terminal d'une séquence (Fu, 2010; Angov, 2011). Actuellement pour palier à cette problématique deux stratégies existent :

- Il faut optimiser les codons présents dans la séquence du gène d'intérêt. Dans ce cas, il s'agit de modifier au cas par cas les codons pouvant se révéler rares dans la bactérie pour les remplacer par des codons « synonymes » mais présents en plus grande quantité. Ce type d'approche pose une problématique, les conséquences réelles des mutations introduites ne sont pas ou peu connues (Wu *et al.*, 2004).

- Il faut optimiser la quantité d'ARNt disponible au moment de la production de la protéine par l'apport de gènes synthétiques codant pour ces ARNt rares. Cette stratégie a permis de créer certaines souches de bactérie *E. coli* comme les souches de type Rosetta™. L'apport des codons rares s'effectue par la présence, au sein de ces souches, d'un plasmide porteur des gènes codant des ARNt spécifiques de codons rares comme par exemple les plasmides de la série *pRARE*. Cette optimisation permet l'expression de gènes issus d'organismes phylogénétiquement éloignés (Burgess-Brown *et al.*, 2008; Brondyk, 2009).

Les modifications post-traductionnelles représentent un critère important dans le choix d'un hôte d'expression. En effet, si *E. coli* propose beaucoup d'avantages, les capacités de la bactérie à pouvoir réaliser des modifications post-traductionnelles sont assez limitées. Chez *E. coli*, la glycosylation, l'amidation, la myristoylation, la palmitation, la sulfation ou encore l'hydroxylation sont absentes. La plupart de ces modifications sont réalisées au niveau du réticulum endoplasmique ou de l'appareil de Golgi absents d'*E. coli*.

Souches d'Escherichia coli modifiées pour l'expression

Le choix d'une souche est devenu un critère supplémentaire dans la stratégie d'expression. La plupart des souches d'*E. coli* utilisées pour l'expression sont dérivées de deux souches ancestrales la souche K-12 et la souche B. La souche K12 est issue d'un prélèvement réalisé sur un malade en rémission de diphtérie. Son utilisation en laboratoire est liée à sa non virulence et sa facilité à être cultivée. La souche B provient d'une souche isolée en 1942 au cours d'une étude sur les bactériophages. Cette souche a été étudiée dans de nombreux laboratoires et a notamment donné la dérivée BL21 connue pour son utilisation en expression de protéines. Ces deux souches ont été séquencées et se distinguent notamment par la composition en protéines présentes au niveau de leur membrane. La souche B produit plus de porines que la souche K12 augmentant alors la perméabilité de sa membrane. Cette caractéristique lui confère une plus grande sensibilité lors de stress (Han *et al.*, 2012). Le tableau 9 présente les principales souches d'expression ainsi que leurs caractéristiques.

Souche d' <i>Escherichia coli</i>	Dérivée de	Principaux atouts de la souche
AD494	K-12	Mutant <i>trxB</i> , facilite la formation des ponts disulfure dans le cytoplasme
BL21	B834	Déficiente pour les protéases <i>lon</i> et <i>ompT</i>
BL21 <i>trxB</i>	BL21	Mutant <i>trxB</i> , facilite la formation des ponts disulfure dans le cytoplasme, déficiente pour les protéases <i>lon</i> et <i>ompT</i>
BL21 CodonPlus-RIL	BL21	Facilite l'expression de protéines eucaryotes qui contiennent les codons rares d' <i>E.coli</i> : AGG, AGA, AUA, CUA, déficiente pour les protéases <i>lon</i> et <i>ompT</i>
BL21 Codon Plus-RP	BL21	Facilite l'expression de protéines eucaryotes qui contiennent les codons rares d' <i>E.coli</i> : AGG, AGA, CCC, déficiente pour les protéases <i>lon</i> et <i>ompT</i>
BLR	BL21	Mutant <i>recA</i> , stabilisation des répétitions en tandem (lié à la réparation de l'ADN), déficiente pour les protéases <i>lon</i> et <i>ompT</i>
B834	B strain	Auxotrophe pour la méthionine, permet l'utilisation de méthionine marquée ou de sélénométhionine pour les études structurales
C41	BL21	Mutant conçu pour l'expression de protéines membranaires
C43	BL21	Double mutant conçu pour l'expression de protéines membranaires
HMS174	K-12	Mutant <i>recA</i> , stabilisation des répétitions en tandem (lié à la réparation de l'ADN), résistance à la rifampicine
Origami™	K-12	Mutant <i>trxB/gor</i> , facilite la formation des ponts disulfure
Origami™ B	BL21	Mutant <i>trxB/gor</i> , facilite la formation des ponts disulfure, déficiente pour les protéases <i>lon</i> et <i>ompT</i>
Rosetta™	BL21	Facilite l'expression de protéines eucaryotes qui contiennent six codons rares d' <i>E.coli</i> dans le plasmide <i>pRARE</i> : AUA, AGG, AGA, CUA, CCC et GGA, déficiente pour les protéases <i>lon</i> et <i>ompT</i>
Rosetta-Gami™	BL21	Facilite l'expression de protéines eucaryotes qui contiennent six codons rares d' <i>E.coli</i> dans le plasmide <i>pRARE</i> : AUA, AGG, AGA, CUA, CCC et GGA, déficiente pour les protéases <i>lon</i> et <i>ompT</i> , Mutant <i>trxB/gor</i> , facilite efficacement la formation des ponts disulfure
Rosetta™ 2	BL21	Facilite l'expression de protéines eucaryotes qui contiennent sept codons rares d' <i>E.coli</i> dans le plasmide <i>pRARE2</i> : AUA, AGG, AGA, CUA, CCC, CGG et GGA, déficiente pour les protéases <i>lon</i> et <i>ompT</i> .
Star™	BL21	Contient le gène muté <i>rne131</i> permettant une meilleure stabilité des ARN messagers dans le cytoplasme, déficiente pour les protéases <i>lon</i> et <i>ompT</i>
Star™ 2	BL21	Contient le gène muté <i>rne131</i> permet la stabilité des ARN messagers dans le cytoplasme, facilite l'expression de protéines eucaryotes qui contiennent sept codons rares d' <i>E.coli</i> dans le plasmide <i>pRARE2</i> : AUA, AGG, AGA, CUA, CCC, CGG et GGA, déficiente pour les protéases <i>lon</i> et <i>ompT</i> .

Tableau 9: Principales souches d'expression d'*E. coli*

Trois souches d'expression BL21 existent, tout d'abord la souche de base, puis la souche « DE3 » permettant l'expression de l'ARN polymérase T7 (donc la transcription du gène présent sur le vecteur d'expression) et enfin la version « DE3 pLysS » permettant de limiter l'expression basale. L'attention doit être portée pour le choix de la souche et sa compatibilité avec le vecteur d'expression. La mise en culture de chaque souche nécessite l'usage d'un milieu de culture adapté. Si la température a démontré des effets sur le métabolisme de la bactérie, les nutriments présents dans le milieu de culture sont également un facteur de stress possible pour la cellule. La disponibilité des sources de carbones et d'azote a un rôle important tout comme la régulation du taux d'oxygène lors des cultures à haute densité de cellules (Tripathi, 2009). La bactérie *E. coli* nécessite enfin la prise en compte de la masse moléculaire de la protéine à exprimer. *E. coli* se révèle, en effet, moins efficace pour les protéines de masse moléculaire supérieure à 50 kDa.

3.2.2. Hôte procaryote : *Bacillus* sp.

Un des avantages du genre *Bacillus*, bactérie Gram +, est la possibilité de sécréter les protéines produites dans le milieu de culture, évitant ainsi les procédés de lyse de cellules. De plus,

contrairement à *E. coli*, ces bactéries ne produisent pas de lipopolysaccharides. Elles ont une excellente capacité de production en utilisant des milieux et des conditions de culture peu coûteuses et simples. De nombreuses optimisations réalisées sur cette bactérie ont permis son utilisation dans les productions industrielles et pharmaceutiques (Westers *et al.*, 2004).

Bacillus sp. présente quelques désavantages, notamment la difficulté à conserver le vecteur d'expression et une importante fragilité lors des processus de transformation pour certaines espèces (Troeschel *et al.*, 2012). La formation des ponts disulfures et la sécrétion de protéases endogènes peuvent de plus provoquer l'altération des protéines recombinantes produites.

Les principales espèces utilisées sont *B. subtilis*, *B. megaterium*, *B. licheniformis* ou *B. brevis*. Dans le cadre d'expression homologue, les bactéries *Bacillus* sp. sont utilisées notamment pour des enzymes du type protéase ou amylase (Demain and Vaishnav, 2009). Pour ce type d'enzymes *Bacillus megaterium* est privilégiée car elle bénéficie de caractéristiques avantageuses. Elle présente peu d'activité protéase, une bonne stabilité du plasmide et une capacité à se développer sur une grande variété de substrats (Meinhardt *et al.*, 1989).

Bacillus subtilis est le second hôte d'expression procaryote le plus utilisé après la bactérie *E. coli*. La présence de codons rares est plus faible que chez *E. coli* (Terpe, 2006). Cette bactérie peut également être transformée par un grand nombre de plasmides et de bactériophages (Yin *et al.*, 2007). La création de souches mutées et/ou le changement du milieu de culture ont permis d'augmenter le niveau d'expression hétérologue (Wong *et al.*, 1986). De nouvelles applications sont à l'étude car son protéome vient d'être publié (Becher *et al.*, 2011).

3.2.3. Hôtes eucaryotes : les levures et les champignons

Généralités

La production hétérologue de protéines dans des hôtes eucaryotes s'est développée principalement à cause des difficultés rencontrées lors de l'utilisation d'organismes procaryotes pour obtenir certaines protéines recombinantes. Si la protéine à produire nécessite des conditions de repliement ou une modification du type glycosylation, l'utilisation d'un hôte eucaryote est alors nécessaire. L'excellente connaissance d'un point de vue biochimique, physiologique et génétique font de ces organismes des hôtes de choix. C'est particulièrement le cas pour les levures *Saccharomyces cerevisiae*, *Pichia pastoris* et pour le champignon *Aspergillus niger*.

Saccharomyces cerevisiae

Cette levure fait partie des organismes eucaryotes les plus connus et par conséquent le premier organisme eucaryote à avoir été utilisé pour l'expression hétérologue de gènes (Boer *et al.*, 2007). *S. cerevisiae* a permis la production du premier vaccin efficace contre l'hépatite B chez l'homme par expression hétérologue (McAleer *et al.*, 1984).

La protéine produite peut être sécrétée par l'intermédiaire d'un peptide signal "facteur α ". Ce peptide signal est constitué d'une préséquence de 19 acides aminés suivi d'une prorégion de 67 acides aminés. La proséquence possède des sites de glycosylation et un site de clivage pour une endopeptidase permettant de cliver la proséquence (Figure 16). Le fonctionnement de ce système se déroule en trois étapes. Tout d'abord, la préséquence est éliminée dans le reticulum endoplasmique au niveau du peptide signal de clivage. Dans l'appareil de Golgi, l'endopeptidase Kex2 clive la proséquence entre la lysine et l'arginine. Enfin, la protéine ste13 clive la liaison de la proséquence au niveau des répétitions glutamate - alanine. La protéine à sécréter va ensuite subir des glycosylations permettant de suivre la voie de sécrétion (Brake *et al.*, 1984).

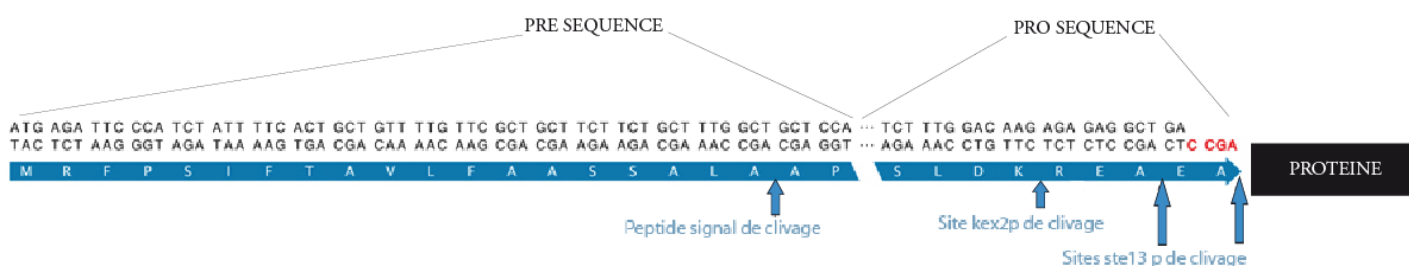


Figure 16: Schématisation du facteur α

La séquence codante de ce peptide signal doit être fusionnée avec la séquence du gène d'intérêt afin d'obtenir un même cadre de lecture et une efficacité dans l'orientation de la protéine vers la voie de sécrétion. Néanmoins, dans certains cas, le clivage entre la séquence et la proséquence n'est pas réalisé ou la production de la protéine recombinée est incomplète. D'autres signaux peptidiques sont donc à l'étude pour corriger ces défauts du facteur α .

La levure *S. cerevisiae* permet l'expression hétérologue de gènes en assurant des modifications post-traductionnelles comme la glycosylation. Cette modification est notamment nécessaire pour l'obtention d'une activité enzymatique, la localisation cellulaire, la stabilité de la protéine ou encore l'interaction entre protéines (Spiro, 2002). L'importante activité de glycosylation de *S. cerevisiae* pose cependant deux problèmes. Tout d'abord, le phénomène d'hyper-glycosylation (ajout d'une centaine de résidus mannose à la place d'une dizaine) des sites de glycosylation modifie la protéine, diminue la quantité produite ainsi que les capacités de sécrétion. De plus, la glycosylation n'est pas identique à

celle réalisée pour les protéines de mammifères, les sites de liaison et les acides aminés visés sont différents. Ce phénomène peut provoquer de mauvaises conformations et des pertes d'activité. Face aux difficultés rencontrées, l'utilisation d'autres levures s'est avérée nécessaire (Demain and Vaishnav, 2009; Valero, 2012).

Pichia pastoris

La levure *Pichia pastoris* est une levure méthylotrophe, utilisant comme seule source de carbone le méthanol. Depuis les 20 dernières années, cette levure a remplacé dans de nombreux domaines *S. cerevisiae*. Son utilisation permet d'atteindre des niveaux d'expression de protéines recombinantes (par exemple l'hirudine) de trois à trente fois supérieurs à ceux obtenus chez *S. cerevisiae* (Sohn et al., 2001). Comme *S. cerevisiae*, *P. pastoris* permet la formation de ponts disulfures et la glycosylation. La production en grande quantité d'une protéine recombinante correctement repliée et fonctionnelle a accru l'utilisation de cette levure. Il est également possible, comme pour *S. cerevisiae*, de sécréter la protéine dans le milieu de culture en utilisant le même peptide signal « facteur α ». Cette capacité de sécrétion permet d'éviter la lyse cellulaire (mécanique ou enzymatique) et de limiter les effets sur les protéines produites. Le nombre de protéines endogènes sécrétées dans le milieu de culture étant limité chez *P. pastoris*, la sécrétion de la protéine recombinante facilite aussi sa purification. Il est également possible d'intégrer la séquence d'intérêt directement dans le génome de cette levure, assurant ainsi la stabilité de l'expression. *Pichia pastoris* peut être mise en culture en fermenteur et le gain de productivité est important avec des densités de cellules de 200 g/L (Heyland et al., 2010). La production de protéines dans *P. pastoris* est en général placée sous le contrôle du promoteur AOX1 permettant d'utiliser le méthanol comme agent d'induction (Celik and Calik, 2012). Il est à noter que les protéines produites peuvent subir une étape de protéolyse, principalement due aux protéases endogènes. Ces protéases semblent être la conséquence du métabolisme du méthanol lors des phases fermentatives en cours de production de protéines recombinantes (Sinha et al., 2005). Ceci implique donc un contrôle précis des temps et des paramètres d'induction.

Yarrowia lipolytica

Y. lipolytica est une levure hétérotrophe et dimorphique, son usage est complémentaire de celui de *P. pastoris*. Cette levure peut notamment produire avec qualité des protéines hétérologues même de masse moléculaire importante. De la même manière que *P. pastoris*, cette production peut se faire en grande quantité. C'est le premier organisme de ce type à avoir été muté pour prévenir l'hyperglycosylation et ainsi se rapprocher de la voie de glycosylation humaine pour les protéines

produites (Song *et al.*, 2007). La majorité des productions de protéines hétérologues par *Y. lipolytica* est réalisée sous le contrôle du promoteur XPR2 qui contrôle normalement la production de la protéase alcaline extracellulaire. Cependant, ce promoteur constitutif a un impact non négligeable sur la croissance de *Y. lipolytica*. D'autres promoteurs constitutifs tels que le promoteur de l'acyl-CoA thiolase (POT1) ou de l'acyl-CoA oxydase (POXx) ont également le même effet (Madzak *et al.*, 2004; Cregg *et al.*, 2009). *Y. lipolytica* ne fermente pas les sucres, son génome a de plus été entièrement séquencé et annoté (Dujon *et al.*, 2004).

Les champignons filamenteux

Certains champignons filamenteux tels qu'*Aspergillus niger* sont capables de sécréter des protéines et des enzymes actives dans un milieu de culture. En industrie, ces enzymes sont utilisées en stabilisant tout d'abord l'organisme producteur et en récoltant en flux continu les protéines produites (Ellaiah *et al.*, 2004). Le gène à exprimer peut être intégré au chromosome par l'intermédiaire d'un plasmide (Kück and Hoff, 2010). L'intégration est réalisée par des mécanismes de recombinaison assurant la stabilité de la transformation. Cependant le grand nombre de copies de plasmides peut aboutir à des intégrations sur des sites non homologues et provoquer des mutations dans les souches d'expression (Lubertozzi and Keasling, 2009). La particularité des champignons filamenteux tient au niveau de la qualité de production et des modifications post-traductionnelles. Dans le cas de *Trichoderma reesei*, les glycosylations sont réalisées de manières comparables à celles produites dans des cellules de mammifères (Salovuori *et al.*, 1987). Cependant, certains gènes peuvent ne pas être exprimés dans les champignons. La principale raison à cette limitation semble être l'étape de transcription, notamment lorsque le nombre de copies du gène à exprimer atteint un certain seuil. Il n'y a pas de proportionnalité entre le nombre de copies du gène et la quantité de protéine produite. A l'inverse, il existe une corrélation entre la quantité d'ARN messenger et la quantité de protéine produite (Verdoes *et al.*, 1995).

3.2.4. Hôtes eucaryotes supérieurs

Cellules d'insectes

L'hyper-glycosylation et la production de protéases, défauts apparus avec les modèles levures et champignons filamenteux, ont nécessité le choix d'un nouveau type d'hôte ; les cellules d'insectes. Ce modèle cellulaire permet d'améliorer les modifications post-traductionnelles proches de celles trouvées dans les cellules de mammifères. L'expression hétérologue est rendue possible par des baculovirus infectant les cellules d'insectes et servant de vecteur pour intégrer le gène à exprimer. Le

plus couramment utilisé est le virus polyèdre *Autographa californica* ciblant le noyau. Ce baculovirus est pathogène pour les cellules de *Spodoptera frugiperda* qui sont largement utilisées comme hôte de production de protéines recombinantes (Demain and Vaishnav, 2009). Ce système a beaucoup évolué. Au départ, le gène d'intérêt modifié avec des séquences du baculovirus était co-transfecté. Il se retrouvait alors intégré au sein du baculovirus par recombinaison homologue et pouvait être exprimé dans la cellule hôte. Cette technique était très aléatoire. Depuis, cette technique a été améliorée en modifiant les baculovirus. L'étape de sélection de clones positifs a été simplifiée par l'introduction d'une séquence létale si elle ne contient pas le gène à exprimer diminuant ainsi le nombre de clones « faux-positifs ». Ce système permet de produire des protéines dans le cytoplasme et des protéines sécrétées. Les mécanismes de sécrétion nécessitent la présence de signaux peptidiques. Les cellules d'insectes ont l'avantage de reconnaître certains signaux peptidiques identiques à ceux retrouvés dans les cellules de mammifères. Cependant, la possibilité de sécrétion des protéines produites a été longtemps problématique à cause du milieu de culture contenant du serum et compliquant les étapes de purification. L'expression en cellules d'insectes permet la réalisation de ponts disulfures et une plus large gamme de glycosylation. Certaines lignées ont été modifiées pour réaliser des N-glycosylations correspondant aux glycosylations normalement trouvées dans les cellules humaines (Harrison and Jarvis, 2006).

Cellules de mammifères

La production hétérologue en cellules de mammifères s'est largement développée ces dernières années et permet maintenant d'obtenir un niveau de production de quelques grammes de protéines par litre de culture (Morrow, 2007). Dans la littérature, la lignée « Chinese Hamster Ovary » (CHO) et la lignée « Human Embryonic Kidney » (HEK) sont les plus décrites. Les cellules HEK sont majoritairement utilisées dans le cadre d'expression transitoire, tandis que les cellules CHO sont utilisées pour l'expression constitutive (Brondyk, 2009; Geisse and Fux, 2009). L'expression transitoire permet de produire rapidement des protéines recombinantes alors que l'expression constitutive permet d'en obtenir de grandes quantités (Durocher *et al.*, 2002). Dans ce type de cellules, la plupart des protéines produites sont sécrétées. De plus, les modifications post-traductionnelles sont très élaborées. Les capacités de glycosylation sont également parmi les plus variées des cellules disponibles pour l'expression hétérologue mais dépendent du type de cellule et de la composition du milieu de culture utilisés (Jenkins *et al.*, 1996).

3.3. Les principaux vecteurs d'expression

Le vecteur est l'outil de base pour le transfert d'un gène vers l'organisme d'expression. Ce sous paragraphe va présenter les principaux vecteurs utilisés en expression hétérologue chez *E. coli* et *S. cerevisiae*. Une première partie est consacrée aux promoteurs et aux étiquettes protéiques. Une seconde partie illustre les autres éléments constitutifs des vecteurs.

Les promoteurs d'expression dans Escherichia coli

Promoteur	Agent d'induction	Niveau d'expression	Principales caractéristiques
<i>lac</i>	IPTG (0,05-2 mM)	Faible à moyen	Promoteur faible, couteux à l'induction
<i>trc / tac</i>	IPTG (0,05-2 mM)	Modéremment élevé	Promoteur fort, régulation de l'expression, couteux et expression basale élevée
<i>T7 ARN polymérase</i>	IPTG (0,05-2 mM)	Très élevé	Promoteur fort, grande capacité d'expression et de régulation, expression basale (sauf en présence de pLys)
<i>pL promoteur du phage</i>	Changement de température de 30 à 42 °C	Modéremment élevé	Nécessite une souche spéciale, pas d'agent d'induction, expression basale
<i>tetA promoteur / opérateur</i>	Tetracycline (200 µg/L)	Variable de moyen à élevé	Bonne régulation, indépendant du métabolisme ou de la souche, faible expression basale
<i>Promoteur araBAD</i>	L-arabinose (0,001-1 %)	Variable de faible à élevé	Bonne régulation, faible expression basale, agent d'induction peu couteux
<i>Promoteur rhaP_{BAD}</i>	L-rhamnose (0,2 %)	Variable de faible à élevé	Bonne régulation, faible expression basale, agent d'induction couteux

Tableau 10: Principaux promoteurs employés avec *Escherichia coli*.
Les concentrations finales dans le milieu de culture sont indiquées entre parenthèse.

De nombreux systèmes de promoteurs existent pour la bactérie *E. coli*. Ils doivent être sources de peu d'expression basale, être finement régulés et fonctionner quelle que soit la souche bactérienne. L'agent d'induction doit être indépendant des composants du milieu de culture. Pour optimiser l'induction, le promoteur doit être un promoteur fort. Le tableau 10 présente les principaux promoteurs utilisés chez *E. coli*. Les quatre premiers promoteurs sont basés sur le système de régulation de l'opéron-lactose chez la bactérie *E. coli*. Le promoteur *lac*, ainsi que les promoteurs *tac* et *trc*, synthétiques et dérivés du promoteur *lac*, présentent un niveau d'expression faible à modéré. Afin d'obtenir un fort niveau d'expression, le promoteur T7, contrôlant la synthèse de la polymérase T7, a été ajouté. Le gène à exprimer est alors sous la dépendance de ce promoteur lui-même étant sous le contrôle du promoteur *lac*. Dans cette configuration, une expression basale existe mais l'utilisation d'une souche DE3 pLys, décrite précédemment, permet de réduire ce phénomène. Ces quatre premiers promoteurs peuvent être induits par l'isopropyl β -D-1-thiogalactopyranoside (IPTG), un analogue de l'allolactose. Cet analogue non métabolisable par la bactérie, permet de maintenir l'induction en continu car sa concentration reste constante dans le milieu de culture. Les autres promoteurs décrits dans le tableau 10 sont moins utilisés. Il est intéressant de noter le cas du

promoteur pL. En effet, ce dernier ne nécessite pas d'agent d'induction car il est régulé par le répresseur cI857 sensible à la température. Le passage de 30 à 42 °C permet l'inactivation du répresseur et ainsi l'expression d'un gène placé en aval du promoteur pL.

Enfin, il existe des promoteurs se caractérisant sous forme d'opérons qui utilisent la tetracycline, l'arabinose ou le rhamnose comme agent d'induction (Terpe, 2006).

*Les promoteurs d'expression dans *Saccharomyces cerevisiae**

Les différents types de promoteurs permettant l'expression dans la levure *S. cerevisiae* peuvent être constitutifs, régulés ou hétérologues (Tableau 11).

Type de Promoteur	Nom	Protéine encodée	Efficacité	Régulation
Constitutif	<i>ADH1</i>	Alcool deshydrogénase 1	Promoteur élevé	
	<i>PGK1</i>	Phosphoglycérate kinase	Promoteur très élevé	
	<i>ENO</i>	Enolase		
	<i>PYK1</i>	Pyruvate kinase	Promoteur élevé	
Régulé	<i>GAL 1-10</i>	Enzyme du métabolisme du galactose	Promoteur élevé	1000X induit galactose
	<i>ADH2</i>	Alcool deshydrogénase 2	Promoteur fort	100X réprimé glucose
	<i>PHO5</i>	Phosphatase acide	Promoteur fort	200X réprimé phosphate
	<i>MET25</i>	Sulphydrylase O-acétyle homosérine	Promoteur moyen	
	<i>CUP1</i>	Métallothiène du cuivre	Pomoteur moyen	20X induit par Cu ₂₊
Hétérologue	<i>CamV</i>	promoteur 35S du virus de la mosaïque du chou-fleur		
	<i>GRE</i>	Récepteur aux glucocorticoïdes		
	<i>ARE</i>	Récepteur aux androgènes		

Tableau 11: Promoteurs utilisés avec la levure *Saccharomyces cerevisiae*

Les promoteurs constitutifs sont souvent dérivés de gènes des voies de dégradation des sucres. C'est le cas notamment des promoteurs ADH1 ou PGK1 qui ont été élaborés au début de la génétique des levures (Boer *et al.*, 2007). Les promoteurs régulables peuvent être contrôlés plus simplement par la disponibilité ou non d'un nutriment dans le milieu de culture. Le nutriment peut autant avoir un effet répresseur qu'un effet inducteur. L'expression peut, dans ce cas, être déclenchée lorsque la densité cellulaire est suffisante. Certains promoteurs provenant de *S. cerevisiae* sont également utilisés par d'autres systèmes d'expression du type levure (PHO5, CUP1, GAL1, GAL7, GAL10 et PGK). Le système d'induction au galactose (GAL1-10) est l'un des mieux connus pour les systèmes d'expression eucaryotes (Johnston and Davis, 1984). Les promoteurs hétérologues sont quant à eux moins efficaces et leur fonctionnement est plus aléatoire.

Les étiquettes et protéines de fusion

Dans la levure comme dans bactérie, l'expression hétérologue autorise l'ajout d'une étiquette peptidique fusionnée à la protéine d'intérêt. Ce sont des éléments placés en position N- ou C-terminal de la protéine à exprimer. Il s'agit de protéines ou de courts peptides qui peuvent avoir plusieurs rôles :

- Favoriser l'expression du gène d'intérêt en fusionnant une protéine hautement exprimée en amont de celle souhaitée.
- Favoriser la solubilité de la protéine en ajoutant une séquence peptidique en N- terminal.
- Favoriser la détection de la protéine recombinante par la présence d'une séquence « épitope » ou d'une protéine connue. Cette séquence permet l'utilisation d'un anticorps spécifique. C'est une technique également valable pour certaines étiquettes de purification. La détection peut également être due à une particule de fusion fluorescente (exemple la GFP).
- Favoriser les conditions de purification en utilisant des étiquettes permettant l'immobilisation de la protéine fusionnée à cette étiquette sur des résines spécifiquement préparées.

Dans certains vecteurs, un site de clivage protéique est présent permettant de supprimer la protéine ou l'étiquette fusionnée. La protéine d'intérêt peut alors être sous forme native après avoir été purifiée par affinité ou isolée sous forme soluble. La présence d'une étiquette nécessite l'ajout d'un codon d'initiation et de terminaison au niveau du vecteur. Le codon d'initiation le plus utilisé demeure le codon ATG. Le codon CTG est utilisé dans 8 % des cas. Pour le codon terminateur, trois codons existent mais le codon TAA est privilégié. Les deux autres codons TAG et TGA peuvent en effet coder respectivement la pyrrolysine et la sélénocystéine. L'efficacité du codon stop est assurée en plaçant 2 ou 3 codons stop en série. De nombreux vecteurs d'expression sont composés d'au moins une étiquette. L'étiquette poly-histidine est la plus couramment utilisée. Elle est compatible avec les systèmes de chromatographie par affinité possédant des ions métalliques immobilisés (connus sous le nom d'IMAC). Sa petite taille favorise cette étape de purification (Peti and Page, 2007). Une seconde étiquette est également utilisée, il s'agit de l'étiquette « Glutathione-S-Transférase ». Cette étiquette de 27 kDa est moins privilégiée car le temps d'interaction avec la résine glutathione-sépharose peut être long (en fonction de la quantité d'extrait protéique).

*Les principaux vecteurs d'expression utilisés avec *Escherichia coli**

Les vecteurs d'expression dans *E. coli* sont généralement construits autour de plusieurs éléments essentiels. La figure 17 présente la conformation de base d'un vecteur d'expression. Chaque type de vecteur a sa conformation type mais des points communs existent. Il faut que chaque vecteur contienne une origine de réplication pour contrôler le nombre de copies du vecteur d'expression. La présence d'un promoteur, dont les caractéristiques ont été décrites dans le paragraphe précédent, permet l'initiation de la transcription. Il est placé 10 à 100 nucléotides avant le site de fixation du ribosome. Le site de clonage multiple (MCS) est l'élément indispensable pour intégrer le gène à exprimer au sein du vecteur d'expression. Sa configuration peut être une succession de sites de restriction mais également un site de clonage unique spécifique de kits commerciaux. L'efficacité de la traduction est dépendante de la position de la séquence Shine-Dalgarno (SD) par rapport à la position du codon d'initiation. La séquence Shine-Dalgarno 5'-TAAGGAGG-3', est positionnée entre 4 et 44 nucléotides en amont du codon d'initiation. Elle permet l'initiation de la traduction. La composition de la séquence est complémentaire de l'extrémité 3' de l'ARN ribosomal 16s. Un marqueur de sélection peut également être présent sur le vecteur. Il s'agit en général d'un gène de résistance à un antibiotique. Le milieu de culture sera alors complété de l'antibiotique adéquat et permettra ainsi la sélection des bactéries transformées.

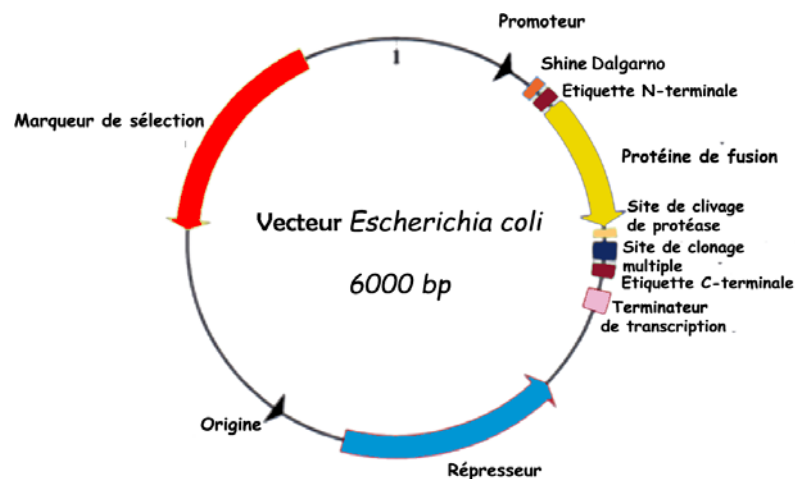


Figure 17: Schéma de base d'un vecteur d'expression d'*Escherichia coli* (European Molecular Biology Laboratory)

Le tableau 12 présente quelques vecteurs d'expression utilisés avec *E. coli*. Les vecteurs choisis dans notre étude apparaissent en grisé. Ce tableau présente l'ensemble des éléments vus précédemment, le type de promoteur et les antibiotiques utilisés. Actuellement, les travaux pour optimiser l'ensemble de ces vecteurs continuent.

Vecteur	Promoteur	Marqueur de sélection	Etiquette ou protéine de fusion	Site de protéase	Vecteur d'origine	Purification d'affinité
pALTER-Ex1	<i>T7</i>	Tetracycline				
pBAD/His	<i>araBAD</i>	Ampicilline	N-Histidine	EntéroKinase	pUC	Oui
pCal-n	<i>T7-lac</i>	Ampicilline	N- « calmodulin-binding-peptide »	Thréonine	ColE1	Oui
pcDNA 2.1	<i>T7</i>	Ampicilline			pUC	
pDUAL	<i>T7-lac</i>	Kanamycine	C- « calmodulin-binding-peptide »	Thréonine	ColE1	Oui
pET15b	<i>T7-lac</i>	Ampicilline	N-Histidine	Thréonine	pBR322	Oui
pET20b(+)	<i>T7</i>	Ampicilline	Séquence signal, C-Histidine		pBR322	Oui
pET28a-c(+)	<i>T7-lac</i>	Kanamycine	N-Histidine, C-Histidine	Thréonine	pBR322	Oui
pET31b(+)	<i>T7-lac</i>	Ampicilline	C-Histidine		pBR322	Oui
pET52b(+)	<i>T7-lac</i>	Ampicilline	N-Strep-Tag, C-Histidine	protéase HRV	pBR322	Oui
pETBlue-3	<i>T7-lac</i>	Ampicilline	N-Histidine, C-histidine	Thréonine / EntéroKinase	pUC	Oui
pGEMEX-1	<i>T7</i>	Ampicilline	gène 10 du phage T7			
pGEX-2T	<i>tac</i>	Ampicilline	« Glutathion S transferase »	Thréonine	pBR322	Oui
pHAT-GFPuv	<i>lac</i>	Ampicilline	N-GFPuv, N-HAT (Histidine-epitope)	EntéroKinase	pUC	Oui
pKK223-3	<i>tac</i>	Ampicilline			pBR322	
pMAL-c2E	<i>tac</i>	Ampicilline	“N-Maltose Binding Protein”	EntéroKinase	ColE1	Oui
pProEXHT (1)	<i>trc</i>	Ampicilline	N-Histidine	TEV		
pRSET	<i>T7</i>	Ampicilline	N-Histidine	EntéroKinase	ColE1	Oui
pSE420	<i>trc</i>	Ampicilline			pUC	
pThioHis	<i>trc</i>	Ampicilline	His-Patch-Trx	EntéroKinase	ColE1	Oui
pTrcHis	<i>trc</i>	Ampicilline	N-Histidine	EntéroKinase	pUC	Oui
pTriEx-1	<i>T7</i>	Ampicilline	C-Histidine		pUC	Oui
pTrxFus	<i>pL</i>	Ampicilline	Thioredoxin	EntéroKinase	ColE1	

Tableau 12: Caractéristiques générales de quelques vecteurs d'expression dans *Escherichia coli*

*Les principaux vecteurs d'expression utilisés avec *S. cerevisiae**

La plupart des vecteurs de la levure sont conçus comme vecteurs « de transfert » entre la bactérie et la levure. Il doit par conséquent à la fois fonctionner dans le système bactérien et le système levure. La plupart de ces vecteurs sont dérivés des plasmides pBR322 et pUC. Ce sont des plasmides bactériens dont une partie des caractéristiques d'origine a été conservée telle que l'origine de réplication ou le marqueur de sélection. Ce sont des vecteurs « hybrides ». Ils permettent d'effectuer les étapes de clonage et de validation dans la bactérie et d'effectuer uniquement l'expression dans la levure. La figure 18 présente l'organisation d'un plasmide d'expression de levure.

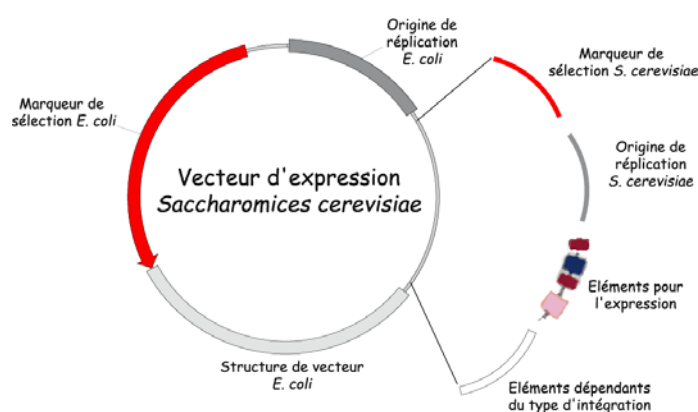


Figure 18: Schématisation d'un vecteur d'expression chez *Saccharomyces cerevisiae*

Les marqueurs de sélection de la levure sont en général un gène codant une enzyme participant à la synthèse d'un acide aminé ou d'un nucléotide. La souche hôte doit être modifiée afin de ne pas posséder le gène en question. Les marqueurs peuvent être soit récessifs (*LEU2*, *TRP1*, *LYS2*, *HIS3*, *URA3*) soit dominants (*CUP1*, *TUN^R*, *KAN^R*). Le tableau 13 présente quelques vecteurs d'expression et leurs éléments caractéristiques.

Vecteur	Promoteur	Marqueur de selection (<i>E. coli</i> / <i>S. cerevisiae</i>)	Vecteur de base	Elements pour l'expression	Elements pour réplication	Elements pour intégration
pAS404	<i>ADH1</i>	Ampicilline / TRP1	pBR322			
pACT2 AD	<i>ADH1</i>	Ampicilline		Terminateur ADH	2μ	
pYC2/NT C	<i>GAL1</i>	Ampicilline / URA3	pBR322	Epitope V5, poly-histidine, terminateur CYC1		CEN6 / ARS4
pAUR123	<i>ADH1</i>	Ampicilline / AureobasidinA				CEN4 / ARS1
<i>pYES2.1</i>	<i>GAL1</i>	Ampicilline / URA3	pUC	Epitope V5, poly-histidine, terminateur CYC1	2μ	
pTEF1	<i>TEF1</i>	Ampicilline / Zeocine	pUC			

Tableau 13: Différents vecteurs d'expression, en grisé le vecteur utilisé lors de cette étude

3.4. Bilan

Le choix des outils pour l'expression hétérologue dépend de nombreux critères qu'il faut définir en fonction des besoins et des moyens à disposition. Le tableau 14 montre les caractéristiques générales à retenir pour chaque hôte.

Caractéristiques	<i>Escherichia coli</i>	Champignons Filamenteux	Levure	Cellules d'insecte	Cellules de mammifères	Cellules de plantes
Temps de croissance	Plusieurs heures à plusieurs jours	Plusieurs jours à une semaine	Plusieurs jours à une semaine	Plusieurs jours à une semaine	Plusieurs semaines	Plusieurs mois
Coût du milieu de culture	faible à moyen	faible à moyen	faible à moyen	élevé	élevé	moyen à élevé
Taux d'expression	faible à élevé	faible à élevé	faible à élevé	faible à élevé	faible à élevé	faible
Evaluation de la méthode	facile	difficile	facile	difficile	difficile	moyen
Capacités de sécrétion	oui dans le périplasme	oui dans le milieu	oui dans le milieu	oui dans le milieu	oui dans le milieu	oui dans le milieu
Modifications post-traductionnelles						
Repliement de protéines	Repliement nécessaire	Repliement pouvant être nécessaire	Repliement pouvant être nécessaire	Repliment assuré	Repliement assuré	Repliement assuré
N-glycosylation	Non	Oui, similaire à ceux des mammifères, pas d'acide sialique, incompatible avec certaines protéines humaines	Oui, élevée en mannose, pas d'acide sialique, incompatible avec certaines protéines humaines	Oui glycosylation complexe, pas d'acide sialique, incompatible avec certaines protéines humaines	Oui glycosylation complexe, incompatible avec certaines protéines humaines	Oui glycosylation complexe, pas d'acide sialique, incompatible avec certaines protéines humaines
O-glycosylation, O-phosphorylation, O-acetylation, O-acylation	Non	Oui	Oui	Oui	Oui	Oui

Tableau 14: Principaux critères de choix d'un organisme d'expression

Le tableau 14 intègre un système d'expression en cellules de plante. Les organismes végétaux sont peu privilégiés pour l'expression hétérologue car le temps de génération est long. De plus, les systèmes de production existant montrent un faible taux d'expression. Néanmoins, les organismes végétaux sont capables de réaliser la plupart des modifications et des repliements nécessaires à une expression hétérologue. Récemment, un système d'expression a été développé permettant d'associer la possibilité de culture à haute densité de cellules avec les qualités d'expression d'organismes végétaux. Ce système d'expression est réalisé sur l'algue *Chlamydomonas reinhardtii* (Griesbeck *et al.*, 2006). Les résultats

sont plus qu'encourageants pour l'expression hétérologue dans cet organisme et notamment pour l'obtention d'enzymes déjà caractérisées (Lauersen *et al.*, 2013).

3.5.L'expression hétérologue d'enzymes lipolytiques

L'expression hétérologue d'enzymes lipolytiques existe dans les laboratoires et industries depuis de nombreuses années. Deux lipases ont été les premières enzymes lipolytiques clonées dans *E. coli*, la lipase linguale du rat et la lipase de *Staphylococcus hyicus* (Docherty *et al.*, 1985; Götz *et al.*, 1985). Ce sous-chapitre présente les études de différentes enzymes lipolytiques à travers différents hôtes d'expression.

Expression hétérologue de lipases

Une revue bibliographique récente a montré que les principaux organismes utilisés pour l'expression hétérologue de lipases sont *Escherichia coli*, *Saccharomyces cerevisiae* et *Pichia pastoris* (Valero, 2012). Le tableau 15 regroupe différentes études sur ces trois hôtes.

Enzyme	Hôte d'expression	Hôte d'origine	Commentaires	Référence
Lipase B	<i>Escherichia coli</i>	<i>Candida antarctica</i>	Expression fonctionnelle	(Blank <i>et al.</i> , 2006)
Lipase B	<i>Escherichia coli</i>	<i>Candida antarctica</i>	Ponts disulfures et repliement	(Xu <i>et al.</i> , 2008)
Lipase	<i>Escherichia coli</i>	<i>Pleurotus sapidus</i>	Lipase extracellulaire, activité avec étiquette	(Krügener <i>et al.</i> , 2009)
Lipase	<i>Escherichia coli</i>	<i>Pseudomonas</i> sp.	Lipase, repliement in vitro	(Akbari <i>et al.</i> , 2010)
Lipase	<i>Escherichia coli</i>	<i>Bacillus subtilis</i>	Lipase extrêmophile	(Shi <i>et al.</i> , 2010)
Lipase	<i>Escherichia coli</i>	<i>Serratia marcescens</i>	Lipase, repliement in vitro	(Li <i>et al.</i> , 2011)
Lipase	<i>Escherichia coli</i>	Métagénomique	Caractérisation complète	(Glogauer <i>et al.</i> , 2011)
Lipase Acyltransferase	<i>Saccharomyces cerevisiae</i>	<i>Candida parapsilosis</i>	Plusieurs séquences codantes	(Neugnot <i>et al.</i> , 2002)
Lipase	<i>Saccharomyces cerevisiae</i>	<i>Yarrowia lipolytica</i>	Lip2, induction par les lipides	(Shockey <i>et al.</i> , 2011)
Lipase B	<i>Pichia pastoris</i> + <i>Escherichia coli</i>	<i>Candida antarctica</i>	Optimisation de l'expression	(Larsen <i>et al.</i> , 2008)
Lipases	<i>Pichia pastoris</i>	<i>Pseudomonas</i> sp.	Métagénomique	(Zhang, 2009)

Tableau 15: Présentation de quelques études de lipases par expression hétérologue

Certaines levures font partie des sources de lipases bien connues. *Candida antarctica* produit par exemple la lipase B. Cette lipase a été étudiée à de nombreuses reprises par expression hétérologue. En 2006, Blank *et al.* se sont intéressés à l'expression de plusieurs mutants de cette lipase. Les auteurs ont d'une part montré le grand intérêt de l'utilisation de différents organismes d'expression, et d'autre part que la bactérie *E. coli* permettait un gain de temps pour tester de nombreuses variétés de mutants. C'est la première étude présentant la production d'une lipase B fonctionnelle après expression dans la bactérie. Ce cas est particulier car la lipase B est normalement excrétée et nécessite

la formation de ponts disulfures. L'expression a donc été réalisée dans le périplasme d'*E. coli* de manière à faciliter la formation de ces ponts. La même étude a menée une production de l'enzyme à une température de 25 °C. Dans ces conditions, l'enzyme a présenté une activité enzymatique comparable à la lipase B produite dans *Aspergillus oryzae*. Afin d'optimiser la formation de ponts disulfures tout en conservant l'enzyme sous sa forme active, Xu *et al.* (2008) ont testé différentes souches d'expression ainsi que différents vecteurs. L'étude est basée sur l'expression de la lipase B de *Pseudozyma antarctica*, un basidiomycète, en évaluant l'impact de l'association de différentes étiquettes et de différents facteurs de repliement. L'expression est comparée entre la souche BL21 et la souche Origami. L'ajout d'une étiquette *DsbA* (étiquette pouvant à la fois permettre la purification par affinité et la production de ponts disulfures) permet l'obtention d'une lipase B plus soluble avec un meilleur repliement, le rôle précis de l'étiquette reste à déterminer. La formation des ponts disulfures (grâce à la souche Origami) permet l'obtention d'une meilleure activité. Dans une autre étude, Krügener *et al.* (2009) ont montré la production de la lipase de *Pleurotus sapidus* sous forme fonctionnelle et non dégradée. Ce résultat a été obtenu pour la lipase produite avec une étiquette poly-histidine. En absence de l'étiquette, la lipase s'est révélée non active. Cette étude a également permis d'explorer les mécanismes de repliement in vitro.

La majorité des expressions hétérologues de lipases est réalisée dans la bactérie *E. coli* et notamment lorsque l'organisme d'origine est une bactérie. La lipase de la souche FS1403 de *Bacillus subtilis* a ainsi été clonée dans *E. coli*. La particularité est ici de produire en condition classique une lipase dont l'origine biologique se situe dans un cratère volcanique. La lipase a été exprimée en utilisant un plasmide de la série pET et une souche BL21. L'expression de cette lipase sous forme active n'a pas nécessité de conditions particulières et s'est déroulée à 37 °C (Shi *et al.*, 2010).

L'étude de la lipase de *Candida parapsilosis* a été particulière. En effet, l'analyse a montré deux cadres ouverts de lecture, chacun porteur du pentapeptide consensus. L'expression hétérologue de ces deux séquences a permis de mettre en évidence celle qui permettait d'obtenir l'enzyme fonctionnelle. Cette étude a utilisé la levure *S. cerevisiae* (Neugnot *et al.*, 2002). La levure *S. cerevisiae* est également utilisée dans de nombreux processus industriels, notamment en fermenteur pour produire des métabolites ou des protéines d'intérêt. Cependant, les triglycérides sont un substrat que la levure ne peut pas utiliser. Récemment, une étude a donc montré l'expression hétérologue de la lipase 2 de *Yarrowia lipolytica* dans *S. cerevisiae*. L'expression a nécessité l'optimisation des codons mais a permis à cette levure de convertir des substrats lipidiques (Shockey *et al.*, 2011).

La levure *Pichia pastoris* est enfin également utilisée pour la production de lipases. Une étude portant sur la lipase B, s'est intéressée à l'impact de l'étape de glycosylation. L'étude a comparé l'expression de cette enzyme dans la levure et dans *E. coli*. Les résultats montrent d'une part que la glycosylation ne permet pas une meilleure activité spécifique et d'autre part que la production rapportée au volume de culture est plus importante chez la levure. De plus, la présence de codons rares chez la bactérie semble être le facteur limitant. L'étape de traduction constitue la limitation majeure de la production dans la bactérie. Toutefois, pour obtenir une meilleure activité spécifique, les paramètres de culture au moment de l'induction semblent essentiels pour les deux hôtes (Larsen *et al.*, 2008).

Expression hétérologue de carboxylestérases

Les carboxylestérases ont été au même titre que les lipases exprimées dans différents hôtes (Tableau 16). Comme dans le cas de certaines lipases, un hôte procaryote (*E. coli*) peut être utilisé pour l'expression d'une carboxylestérase d'eucaryote (porc). De la même manière, la levure *S. cerevisiae* a été utilisée pour l'expression d'une carboxylestérase de bactérie. Si globalement le principal objectif de ces études est la caractérisation biochimique, certaines études cherchent également à illustrer les distinctions entre les lipases et les carboxylestérases.

Enzyme	Hôte d'expression	Hôte d'origine	Commentaires	Référence
Carboxylestérase	<i>Escherichia coli</i>	<i>Bacillus licheniformis</i>	Similarités avec des lipases	(Alvarez-Macarie <i>et al.</i> , 1999)
Carboxylestérase	<i>Escherichia coli</i>	<i>Thermoanaerobacter tengcongensis</i>	Similarités avec des lipases	(Zhang <i>et al.</i> , 2003)
Carboxylestérase	<i>Escherichia coli</i>	<i>Bacillus coagulans</i>	Motif GX SXG	(Mnisi <i>et al.</i> , 2005)
Carboxylestérase	<i>Escherichia coli</i>	<i>Sus domesticus</i>	Origami-molécules chaperonnes	(Bottcher <i>et al.</i> , 2007)
Carboxylestérase	<i>Escherichia coli</i>	<i>Fervidobacterium nodosum</i>	Motif GDSL	(Yu <i>et al.</i> , 2010)
Carboxylestérase	<i>Escherichia coli</i>	<i>Malus domestica</i>	Motif GX SXG	(Souleyre <i>et al.</i> , 2011)
Carboxylestérase	<i>Saccharomyces cerevisiae</i>	<i>Thermus thermophilus</i>	Expression fonctionnelle	(Lopez-Lopez <i>et al.</i> , 2010)
Carboxylestérase	<i>Pichia pastoris</i>	<i>Sus domesticus</i>	Nouvelle isoenzyme, nouvel hôte d'expression	(Hermann <i>et al.</i> , 2008)

Tableau 16: Présentation de quelques études de carboxylestérases par expression hétérologue

Ce type d'approche peut être illustré avec un gène issu de *Bacillus licheniformis* dont les premières analyses ont orienté vers l'identification d'une lipase. Ainsi, le site consensus AX SXG caractéristique des lipases de *Bacillus* sp. est bien présent. De plus, la séquence protéique prédictive ne partage pas de similarités avec des séquences de carboxylestérases. Son expression dans *E. coli* a cependant permis de caractériser une activité spécifique des chaînes courtes. Cette protéine est donc identifiée comme une carboxylestérase issue probablement de la mutation d'un gène de lipase (Alvarez-Macarie *et al.*, 1999). La vérification biochimique par expression hétérologue de gènes annotés permet ainsi de

valider des gènes susceptibles de coder des lipases. Dans une autre étude faite par Zhang *et al.* (2003), la séquence protéique présente des similitudes entre lipase et carboxylestérase. Seule l'activité carboxylestérase a été caractérisée et a permis d'identifier une carboxylestérase de *Thermoanaerobacter tengcongensis*, organisme se développant à des températures supérieures à 50 °C. Cette distinction a également été montrée dans une étude de Bottcher *et al.* (2007) pour la gamma-isoenzyme carboxylestérase de foie de porc (γ -PLE). Plusieurs essais ont été réalisés pour exprimer cette enzyme dans la bactérie *E. coli* ainsi que dans la levure *P. pastoris*. Dans le premier hôte, aucune expression n'a été détectée, tandis que dans la levure une faible quantité a été produite. Cette étude est caractéristique de l'utilisation de souches d'expression particulières d'*E. coli* pour réaliser l'expression de protéines eucaryotes. La γ -PLE a ainsi été exprimée dans les souches dérivées de la souche BL21, Origami™, Rosetta™ et Rosetta-gami™, essentiellement en corps d'inclusion. Les protéines recombinantes solubles se sont avérées inactives. Pour réaliser l'expression fonctionnelle de cette carboxylestérase, une stratégie complémentaire a été mise en œuvre en utilisant une molécule chaperonne. Cette étude a également montré l'incapacité de la bactérie à réaliser des glycosylations n'empêchant pas la protéine d'être active. Curieusement, l'expression hétérologue des carboxylestérases semble privilégiée dans la bactérie *E. coli*. La caractérisation de l'activité carboxylestérase du gène MdCXE1 de la pomme a également été réalisée et permis l'obtention d'une protéine fonctionnelle (Souleyre *et al.*, 2011).

L'expression hétérologue est choisie dans certains cas lorsque l'obtention de la protéine dans l'hôte d'origine est coûteuse ou requiert des conditions complexes. C'est le cas d'une des carboxylestérases de *Thermus thermophilus*, bactérie Gram-négative dont la température de développement est proche de 65 °C. La production dans *S. cerevisiae* a été privilégiée car elle présente un système de sécrétion. La carboxylestérase produite est fonctionnelle. Cependant, si le système de signalisation de la protéine semble fonctionner, une large quantité de la protéine est retenue au niveau du périplasme (80%). De plus, une étape de glycosylation est nécessaire (Lopez-Lopez *et al.*, 2010).

L'utilisation de *Pichia pastoris* comme hôte d'expression a permis de produire l'APLE, variant de la γ -PLE. Néanmoins au cours de cette étude, il a été démontré que les protéines APLE semblaient former au minimum un trimère et que la voie de sécrétion de la levure n'était pas forcément adaptée pour cette taille de molécule. Aucune activité extracellulaire n'a été caractérisée, la protéine semble retenue dans l'espace périplasmique. Une fois purifiée et isolée du périplasme, l'enzyme est active à 64U / mg sur le p -nitrophenol acétate (Hermann *et al.*, 2008).

Expression hétérologue de thioestérases

L'expression d'Acyl-ACP thioestérase est depuis une dizaine d'année majoritairement réalisée dans la bactérie *E. coli* (63 % des cas) (Source : Web of Knowledge). En effet, la connaissance du métabolisme de la bactérie permet, dans un certain nombre de cas, de non seulement exprimer la protéine mais également d'évaluer sa fonctionnalité directement sur la bactérie (Tableau 17).

Enzyme	Hôte d'expression	Hôte d'origine	Commentaires	Référence
Thioestérase	<i>Escherichia coli</i>	<i>Brassica juncea</i>	Quatre ADNc caractérisés	(Jha <i>et al.</i> , 2010)
Thioestérase	<i>Escherichia coli</i>	<i>Phaeodactylum tricornutum</i>	Rosetta™, K27	(Gong <i>et al.</i> , 2011)
Thioestérase	<i>Escherichia coli</i>	Diverses plantes	Comparatif	(Zhang <i>et al.</i> , 2011)
Thioestérase	<i>Escherichia coli</i>	<i>Arachis hypogaea</i>	Caractérisation	(Chen <i>et al.</i> , 2012)
Thioestérase	<i>Escherichia coli</i>	<i>Acinetobacter baylyi</i>	Production d'acides gras	(Zheng <i>et al.</i> , 2012)

Tableau 17: Présentation de quelques études de thioestérases par expression hétérologue

Les enzymes caractérisées sont de différentes origines. L'activité d'enrichissement est *in fine* recherchée, les organismes d'origine sont le plus souvent des plantes ou des microorganismes accumulant des lipides. Le plant de moutarde, par exemple, produit une série de gènes qui peuvent coder des protéines à activité thioestérase. Quatre ADN complémentaires ont donc été isolés et clonés dans la bactérie *E. coli*. Quatre isoformes d'enzymes du type « Fatty acyl-ACP Thioestérase » ont ainsi été produites. L'expression est réalisée dans des souches modifiées pour la β -oxydation. L'expression de chaque isoforme dans *E. coli*. a entraîné une modification du profil d'acides gras des lipides totaux de la bactérie, traduisant ainsi une expression fonctionnelle (Jha *et al.*, 2010). Dans une autre étude, l'expression de thioestérases provenant de diverses plantes a été réalisée. L'étude illustre notamment le rôle possible d'une thioestérase exprimée de manière hétérologue. Son intervention semble se situer dans le processus de production des lipides. Si l'enzyme est active, son intégration s'effectue dans le cycle de synthèse (Figure 19). Le profil en acides gras des lipides totaux de la bactérie est alors fonction des spécificités de l'enzyme (Zhang *et al.*, 2011). Ce type d'approche a également été testé sur une thioestérase d'*Acinetobacter baylyi*. La bactérie utilisée est alors une souche classique de surexpression BL21(DE3). L'expression de cette enzyme a permis non seulement de modifier le profil mais également d'augmenter globalement l'ensemble des acides gras. Cette étude a également montré l'importance de la composition du milieu de culture ainsi que le choix du promoteur pour l'optimisation de la quantité d'enzyme produite, privilégiant ainsi le système T7 (Zheng *et al.*, 2012).

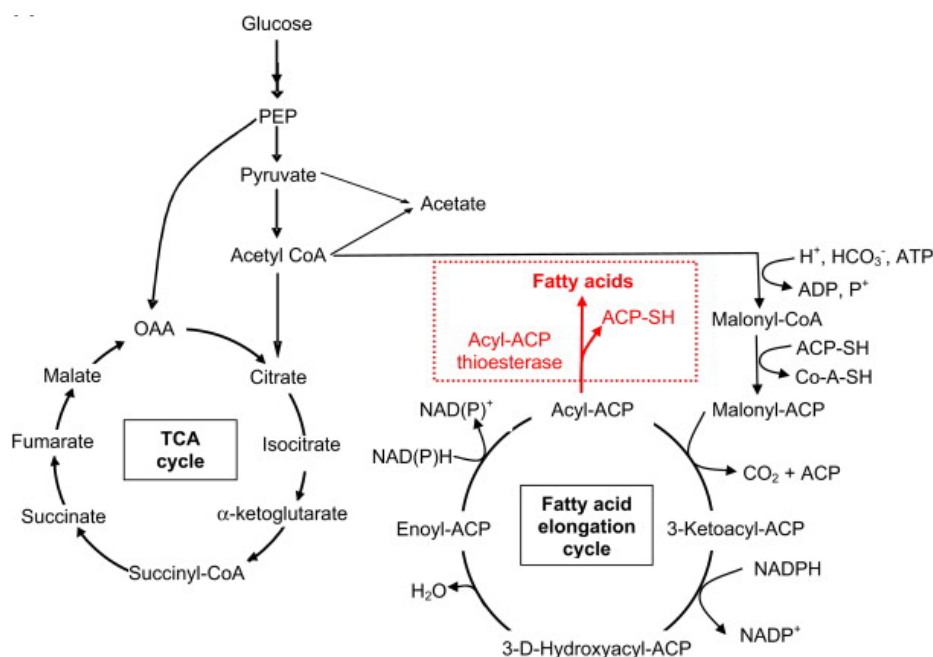


Figure 19: Illustration de la place des thioestérases dans la synthèse des acides gras (Zhang *et al.*, 2011)

Expression hétérologue d'autres activités lipolytiques

Ce dernier paragraphe présente d'autres enzymes à activité lipolytique (Tableau 18).

Enzyme	Hôte d'expression	Hôte d'origine	Commentaires	Référence
Cutinase like	<i>Escherichia coli</i>	<i>Mycobacterium tuberculosis</i>	CULP 1-7	(West <i>et al.</i> , 2009)
Cutinase like	<i>Escherichia coli</i>	<i>Mycobacterium tuberculosis</i>	CULP 1 & 4	(Schue <i>et al.</i> , 2010)
Ester Hydrolase	<i>Escherichia coli</i>	Mycobacteriophage Ms6	Activité lipolytique	(Gil <i>et al.</i> , 2008)

Tableau 18: Présentation des autres activités lipolytiques

Les cutinases présentes au sein d'organismes pathogènes comme *Mycobacterium tuberculosis* montrent un intérêt croissant. Une étude sur l'expression hétérologue de cutinases issues de *M. tuberculosis* chez *E. coli* a comparé l'activité de cette enzyme avec une cutinase vraie de *Thermobifida fusca* utilisée comme témoin positif. Seule une activité de type « Cutinase Like Protein » (CULP) a été observée (West *et al.*, 2009). Ces auteurs ont voulu valider les acides aminés impliqués dans l'activité enzymatique. Ils ont testé différentes mutations du site catalytique par remplacement successif de chacun des acides aminés de la triade catalytique par une alanine. Aucune activité n'a alors été détectée.

Cette démarche expérimentale a été optimisée pour l'étude de deux de ces CULPs. L'obtention de ces enzymes a permis de mieux caractériser leurs cibles physiologiques et les possibles clés de leur

virulence. Cette étude par mutation et par modélisation moléculaire a permis l'identification des acides aminés impliqués dans la triade catalytique de ces CULPs (Schue *et al.*, 2010).

Enfin, dans le cas du mycobactériophage Ms6, l'activité lipolytique est recherchée sur la base de la présence du pentapeptide GX SXG. Ce pentapeptide est présent au niveau de la séquence protéique déduite d'un des gènes codant une protéine du module de lyse du mycobactériophage. La caractérisation de l'activité enzymatique reste donc à réaliser et la séquence codante est alors clonée dans la bactérie *E. coli*. Différentes activités sont recherchées ainsi que la spécificité de la protéine produite. La protéine est active et hydrolyse différentes longueurs de chaînes carbonées, ne permettant pas d'identifier une spécificité. De plus, la présence d'une étiquette de plusieurs histidines ne semble pas affecter l'activité de l'enzyme (Gil *et al.*, 2008).

CHAPITRE II : MATERIEL ET METHODES

CHAPITRE II : MATERIEL ET METHODES

1. Matériels biologiques

1.1. *Saccharomyces cerevisiae*

Dans la première partie du travail de thèse, la souche INVSc1 (MATa his3D1 leu2 trp1-289 ura3-52) de la société Invitrogen™ a été utilisée pour l'obtention de protéines hétérologues dans le modèle levure : *Saccharomyces cerevisiae*. Cette souche a permis de tester les constructions plasmidiques pour l'obtention des deux protéines *IgTeCe* et *IgEst3*.

1.2. *Escherichia coli*

Dans la seconde partie du travail de thèse, le modèle bactérien a été privilégié. Les techniques de clonage ont requis l'usage d'une souche de clonage *E. coli* : TOP10 (F-mcrA Δ(mrr-hsdRMS-mcrBC) φ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(ara-leu) 7697 galU galK rpsL (StrR) endA1 nupG λ-).

L'étude de l'expression d'*IgTeCe* chez *E. coli* a nécessité l'utilisation des souches dédiées à l'expression hétérologue de protéines recombinantes, BL21 Star™ 2 (DE3, F- ompT hsdSB (rB-mB-) gal dcm rne131 pRARE2 (CamR)) et Rosetta™ 2 (DE3, pLysS : F- ompT hsdSB (rB-mB-) gal dcm (DE3) pLysSRARE2 (CamR)). Ces souches sont dérivées du type BL21 et disposent d'un plasmide *pRARE2* porteur de gènes codant pour des ARNt rares chez *E. coli*.

1.3. Plasmides de clonage et vecteurs d'expression

1.3.1. *Saccharomyces cerevisiae*

L'expression chez la levure est basée sur l'utilisation du plasmide *pYES2.1*® (Invitrogen™) (Figure 20). Ce vecteur permet la synthèse de protéines recombinantes dans la souche INVSc1 de la levure *Saccharomyces cerevisiae*.

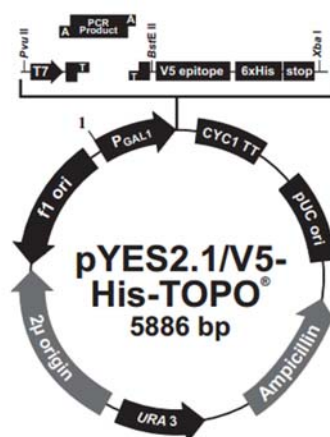


Figure 20: Le vecteur d'expression *pYES2.1* (Invitrogen™)

L'insertion d'un gène au sein de la cassette de clonage est réalisée grâce à la technique "TopoIsomérase" (Figure 21). Le clonage des gènes *IgTeCe* et *IgEst3* dans le plasmide *pYES2.1* a permis d'obtenir les constructions plasmidiques nommées *pYE-IgTeCe* et *pYE-IgEst3*.

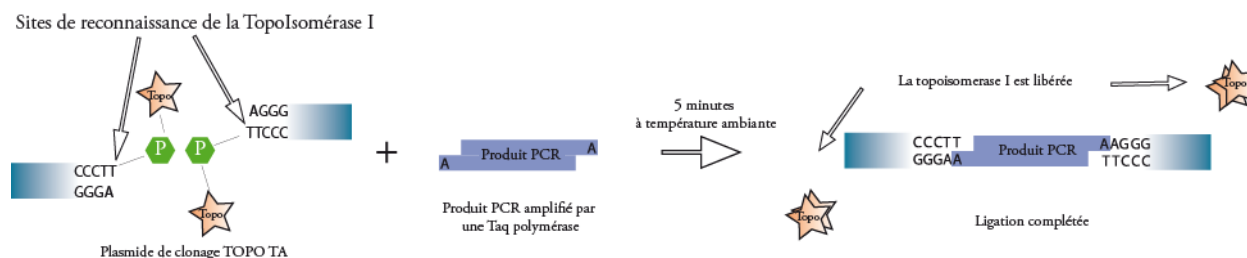


Figure 21: Clonage selon la technique à la Topoisomérase (d'après Invitrogen™)

L'expression est placée sous le contrôle d'un système de répression/induction utilisant le glucose comme répresseur et le galactose comme agent d'induction. Ce plasmide dispose d'un gène codant pour l'uracile permettant une sélection sur un milieu dépourvu de cette base azotée (INVSc1 étant auxotrophe pour l'uracile). La protéine produite à l'aide de ce plasmide est pourvue d'une étiquette poly-histidine à l'extrémité C-terminal.

1.3.2. *Escherichia coli*

Le plasmide de clonage *pJET1.2*

Lors du clonage, le plasmide *pJET1.2* (Figure 22) fourni par la société Fermentas™ a été utilisé.

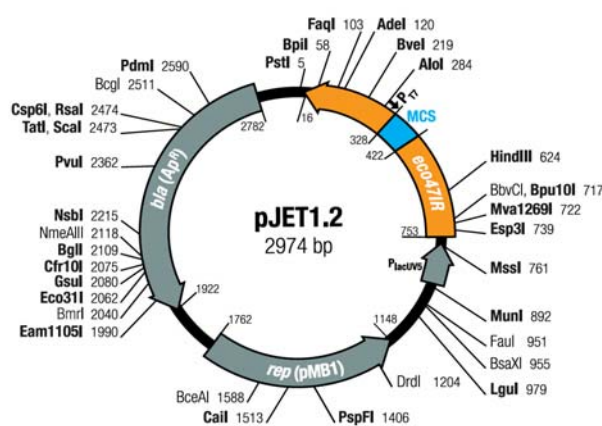


Figure 22: Le plasmide de clonage *pJET1.2* (Fermentas™)

Il est associé à un kit de clonage CloneJet™ provenant de Finnzyme®. Ce plasmide est porteur d'un gène "suicide" dont la séquence contient le site de clonage multiple et le site d'insertion des fragments produits par PCR. Si la ligation est positive et que l'intégrité du gène "suicide" est altérée, le clone sera en mesure de se développer. Dans le cas contraire, aucun clone ne pourra se développer pour être sélectionné.

Les vecteurs d'expression

Le choix de vecteurs d'expression dans l'hôte *E.coli* s'est porté sur les vecteurs de la série “*pET*” de la société Millipore™ : *pET28a+* et *pET52b+*. Ces deux vecteurs utilisent le système d'induction IPTG/T7 compatible avec les souches BL21 choisies. Le vecteur *pET28a+* (Figure 23) est utilisé afin de produire une protéine pourvue d'une ou deux étiquettes poly-histidine. L'insertion du gène à exprimer s'effectue entre les sites de restriction *NcoI* et *NotI* de la cassette de clonage ou entre les sites de restriction *NdeI* et *NotI*.

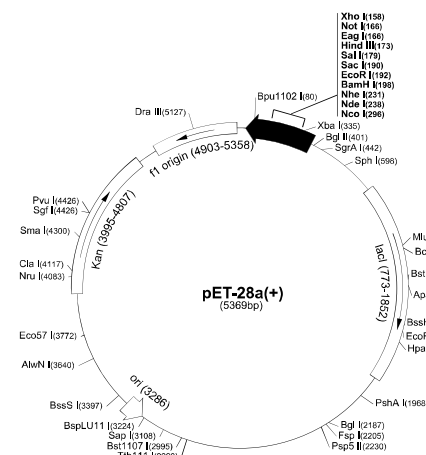


Figure 23: Le vecteur d'expression *pET28a+*

Le vecteur *pET52b+* (Figure 24) est utilisé afin de produire une protéine pourvue d'une étiquette poly-histidine clivable en position C-terminal. L'insertion du gène pourra s'effectuer comme pour le vecteur *pET28a+* au niveau des sites de restriction *NcoI* et *NotI* au sein de la cassette de clonage.

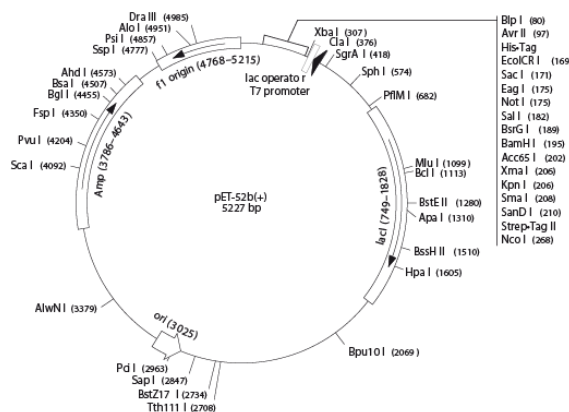


Figure 24: Le vecteur d'expression *pET52b+*

1.4. Milieux de culture

1.4.1. Milieu de culture pour *Saccharomyces cerevisiae*

La culture de la levure s'effectue sur un milieu synthétique minimum supplémenté avec les acides aminés nécessaires, d'un sucre comme source de carbone (glucose ou galactose) et dépourvu de la base azotée uracile. La composition de ce milieu est la suivante : 0,67 % (p/v) d'une base de levure

azotée, 2 % (p/v) de sucre, 0,01 % (p/v) d'adénine, d'arginine, de cystéine, de leucine, de lysine, de thréonine, de tryptophane ; 0,005 % (p/v) d'acide aspartique, d'histidine, d'isoleucine, de méthionine, de phénylalanine, de proline, de sérine, de tyrosine et de valine. Ce milieu sera identifié milieu SC-U. L'ensemble des composants de ce milieu est fourni par la société Sigma-Aldrich™. Dans le cas d'un milieu gélosé, il sera complété par de l'agar-agar (15 g/L).

1.4.2. Milieu de culture pour *Escherichia coli*

Le milieu de culture pour bactérie est un milieu Lysogeny Broth (LB) dans sa variation Lennox comprenant 10 g de peptone, 5 g d'extraits de levure et enfin 5 g de chlorure de sodium par litre de milieu. L'ensemble des composants de ce milieu provient du fournisseur Sigma-Aldrich™. Ce milieu sera celui de base à compléter des antibiotiques nécessaires, et dans le cas d'un milieu gélosé il sera complété par de l'agar-agar (15 g/L).

Chaque vecteur utilisé est porteur d'un gène de résistance à un antibiotique. Ce gène est sous l'influence d'un promoteur. La présence de ce gène de résistance va permettre la sélection des bactéries porteuses de ce plasmide et implicitement du gène véhiculé par le plasmide. Les antibiotiques utilisés lors de ces cultures (milieux liquide ou gélosé) sont indiqués dans le tableau 19.

Antibiotiques fournis par Sigma-Aldrich™	Concentration massique finale dans le milieu	Plasmides utilisés
Ampicilline	100 µg/mL	<i>pJET1.2, pET52b+</i>
Chloramphénicol	34 µg/mL	<i>pRARE2</i>
Kanamycine	30 µg/mL	<i>pET28a+</i>

Tableau 19: Antibiotiques utilisés

1.5. Conditions de culture et de conservation

1.5.1. Suivi de croissance

La densité cellulaire est estimée par turbidimétrie à 595 nm. Les mesures de turbidité sont effectuées par comparaison avec un blanc composé de milieu de culture stérile (LB ou SC-U).

1.5.2. Méthodes de conservation

Conservation avec du glycérol

Les cellules sont mises en culture jusqu'à une turbidimétrie de 0,8. Pour 800 µL de suspension, un volume de 200 µL de glycérol à 85 % est ajouté puis le tube est vortexé avant d'être placé au congélateur à -70°C.

Conservation en boîte de Petri sur milieu gélosé

Après une culture en milieu liquide de 16 heures, l'ensemencement est réalisé en surface par stries à l'aide d'une pipette Pasteur sur un milieu gélosé. Les boîtes de Petri sont ensuite placées à 37°C à l'étuve jusqu'à apparition de colonies le long des stries. Les boîtes sont alors fermées par du parafilm puis conservées à 4°C.

1.5.3. Méthodes de culture de *Saccharomyces cerevisiae*

Les levures transformées possédant le gène à exprimer dans le vecteur d'expression *pYES2.1* sont transférées depuis un milieu de conservation vers le milieu de culture. La préculture s'effectue dans un Erlenmeyer de 500 mL contenant 100 mL de milieu de culture SC-U. Le prélèvement est réalisé à l'aide d'une pipette Pasteur et la préculture est mise sous agitation (170 rpm) dans un incubateur à une température de 30°C pendant 12 heures. A l'issue de la préculture, la turbidité à 595 nm est mesurée.

La culture est alors réalisée dans deux Erlenmeyers de 500 mL contenant chacun 100 mL de milieu SC-U. Chaque Erlenmeyer est inoculé de manière à obtenir une turbidimétrie au départ comprise entre 0,1 et 0,15 à une longueur d'onde de 595 nm. Les cultures sont effectuées en parallèle pour l'essai et pour le contrôle positif, fourni avec le plasmide *pYES2.1*, dans les mêmes conditions que les précultures.

1.5.4. Méthodes de culture d'*Escherichia coli*

E. coli est mis en culture en Erlenmeyer de 250 mL contenant 50 mL de milieu LB en présence d'antibiotiques. La mise en culture s'effectue à partir d'une préculture réalisée dans 50 mL de milieu de culture (LB + antibiotique) à 37°C pendant 12 heures. Le lendemain, la turbidité de la suspension bactérienne est évaluée. Les cultures sont par la suite inoculées et incubées dans les mêmes conditions (turbidité minimum) que celles décrites au paragraphe 1.5.3. Toutes les cultures d'*E. coli* sont placées en incubateur à une température de 37°C et avec une agitation de 175 rpm. Des milieux de culture gélosés en boîtes de Petri sont également utilisés pour la sélection de clones transformés (clonage et sous-clonage).

2. Etude de l'expression chez *Saccharomyces cerevisiae*

2.1. Préambule

L'étude de l'expression chez *S. cerevisiae* est le fruit d'un travail préalable d'isolement et de clonage de gènes réalisé au sein du laboratoire. Le résultat de ces travaux est l'obtention de deux clones contenant *pYE-IgTeCe* et *pYE-IgEst3*. Un séquençage a déjà été réalisé.

Le travail de thèse a été, dans un premier temps, de vérifier l'intégration des gènes. Dans un second temps, l'étude de l'expression des gènes a été entreprise. Les techniques de biologie moléculaire sont issues ou dérivées des protocoles fournis avec les kits commerciaux. Les techniques de base utilisées proviennent de protocoles décrits dans "Molecular cloning, a laboratory manual" (Green *et al.*, 2012).

2.2. Extraction du vecteur *pYES2.1* contenant les gènes présents dans *Saccharomyces cerevisiae*

L'isolement d'ADN plasmidique depuis la levure est effectué en suivant un protocole compatible avec la souche INVSc1 (Hoffman and Winston, 1987). Des cultures de levure de 10 mL en milieu SC-U sont réalisées pendant 16 heures. Une fraction de 2 mL est ensuite transférée dans un microtube puis centrifugée durant 10 min à 4 000 g à une température de 4°C. Le culot obtenu est repris dans un volume de 200 µL de tampon de lyse (2% Triton X-100, 1 % SDS, 100 mM NaCl, 10 mM Tris HCl (pH 8,0), 1 mM EDTA) puis 200 µL de phénol : chloroforme : isopropanol (25 :24 :1). Après ajout d'un volume de 200 µL de billes de verre (lavées à l'acide et de diamètre inférieur à 0,7 mm, fournies par Sigma-Aldrich™), la lyse cellulaire est réalisée par agitation au vortex pendant 2 min. Le tube est centrifugé durant 5 min à 15 000 g à une température de 4°C. La phase aqueuse est ensuite transférée dans un nouveau tube. L'ajout de 2,5 volumes d'éthanol absolu et de 1/10^e de volume d'acétate d'ammonium (4 M) permet la précipitation de l'ADN plasmidique. Le mélange obtenu est incubé à -20°C durant 30 min, puis centrifugé à 15 000 g pendant 10 min à 4°C. Le culot d'ADN est alors récupéré et lavé délicatement avec de l'éthanol 70% puis centrifugé à 15 000 g avant d'être asséché. L'obtention d'une préparation d'ADN plasmidique est complète en reprenant ce culot dans 30 µL d'H₂O stérile. La qualité et la quantité de l'ADN plasmidique sont déterminées par spectrophotométrie à 260 nm et par visualisation sur un gel d'agarose (volume total de dépôt de 10 µL pour 2 µL d'échantillon).

2.3. Amplification par PCR

La réaction de polymérisation en chaîne (PCR), a été mise au point par Mullis *et al.* (1986). Cette technique permet de cibler et d'amplifier *in vitro* une région précise de l'ADN. La région cible est spécifiquement délimitée par un couple d'amorces. Le principe de la PCR est basé sur la répétition d'un cycle précis composé des étapes suivantes :

- Une étape de dénaturation, permettant de séparer les deux brins d'ADN matrice.
- Une étape d'hybridation, permettant l'appariement de l'amorce à la séquence de l'ADN matrice qui lui est complémentaire.
- Une étape d'élongation au cours de laquelle l'ADN polymérase thermostable va synthétiser l'ADN depuis l'extrémité 3'OH de l'amorce dans le sens 5'P vers 3'OH.

A chaque nouveau cycle, les produits d'amplification des cycles précédents servent de matrice à la polymérase, permettant ainsi un doublement du nombre de copies obtenues. La composition du milieu réactionnel (25 µL) et le programme d'amplification sont décrits dans le tableau 20 A et tableau 20 B.

Composition du milieu réactionnel – 20 A
ADN matrice (100 pg d'ADN plasmidique)
0,4 µM de chaque oligonucléotide amorce
200 µM de chaque dNTP
1,5 mM de MgSO ₄
2,5 µL de tampon Vent PCR 10X (New England Biolabs)
1 U de Vent® ADN polymérase (New England Biolabs)

Programme d'amplification du thermocycleur – 20 B	
Dénaturation préliminaire de 3 min à 94 °C	
35 cycles comportant successivement:	Dénaturation de 1 min à 94 °C
	Hybridation de 45 secondes à une température T _m d'hybridation
	Synthèse d'ADN à 72 °C pendant 30 secondes à 70 secondes (selon la taille du fragment)
Synthèse d'ADN à 72 °C pendant 7 min	
Refroidissement de chaque milieu réactionnel à 4°C	

Tableau 20: A)Composition et B)programme pour la réaction PCR

Cette méthode d'amplification par PCR est réalisée dans un premier temps pour amplifier la région contenant les gènes *IgTeCe* ou *IgEst3* dans *pYES2.1*. Elle sera, dans un second temps, utilisée dans d'autres applications (identifier un insert, RT-PCR). Les différentes températures d'hybridation et temps d'élongation nécessaires seront indiqués au fur et à mesure de leurs utilisations. Toutes les amorces sont synthétisées par la société Eurofins MWG.

2.4. Détermination de la présence d'un insert dans le vecteur *pYES2.1*

La détermination de la présence d'un insert avant l'envoi au séquençage a été fait par réaction PCR en choisissant des amorces ciblant le vecteur. Les amorces utilisées sont présentées dans le tableau 21.

Dénomination	Sequence 5' → 3'	Taille en pb	Température d'hybridation °C	Nature de l'amorce
K7EXTRACTFW	GCGAAGCGATGATTTTTGAT	20	52,2	Spécifique de <i>pYES2.1</i>
K7EXTRACTRV	CTTTTCGGTTAGAGCGGATG	20	53,4	Spécifique de <i>pYES2.1</i>

Tableau 21: Amorces utilisées pour déterminer la présence de l'insert

La réaction PCR a été réalisée dans un volume final de 25 µL. La température d'hybridation utilisée est $T_m = 57^\circ\text{C}$ et le temps d'élongation a été fixé à 1 min, 35 cycles ont été nécessaires.

2.5. Electrophorèse et purification des produits amplifiés

Electrophorèse, coloration et visualisation

Les produits PCR ont été systématiquement visualisés et analysés sur gel d'agarose (1,5 % (p/v)) dans un tampon Tris Acetate EDTA 1X (TAE). Cette solution a été préparée à partir d'une solution mère TAE 50X contenant par litre, 242 g Tris Base ($MM = 121,1 \text{ g/mole}$), 57,1 mL d'acide acétique glacial et 100 mL de solution à 0,5 M d'EDTA. La taille des fragments a été déterminée à l'aide du marqueur de taille GeneRuler 1kb DNA Ladder de la société Fermentas™. Les échantillons sont chargés (5/6^e du volume final) avec le tampon 6X DNA Loading Dye (1/6^e du volume final) de Fermentas™. La migration a été réalisée sous une tension constante de 100 V. Afin de visualiser les fragments d'ADN aux rayons UltraViolet (UV), la coloration des bandes est effectuée par trempage dans un bain contenant du bromure d'éthidium (BET). La prise de vue est réalisée sous UV sur un appareil GelDoc™ de BioRad®.

Extraction et purification des produits PCR

Les produits PCR sont séparés sur gel d'agarose avant d'être purifiés. Le gel d'agarose est incubé dans un bain de BET durant un temps inférieur à 5 min puis soumis pour visualisation sur banc UV pour une durée d'exposition inférieure à 1 min et découpé à l'aide d'un scalpel. Ces deux temps d'expositions courts ont permis de prévenir toute altération des fragments d'ADN à récupérer. Une fois découpés, les segments d'agarose contenant les fragments d'ADN sont placés dans des microtubes. La purification est réalisée à l'aide du kit "GenElute™ Agarose Spin Columns" de la société Sigma-Aldrich™ dans les conditions préconisées par le fournisseur et permet d'obtenir un

volume final de 20 à 50 μ L de solution. Celle-ci est stockée à -20°C avant utilisation ou envoi sur plateforme de séquençage.

2.6. Quantification des acides nucléiques

Les échantillons sont quantifiés selon deux méthodes. La première se fait par évaluation sur gel. Deux microlitres des échantillons sont déposés sur gel d'agarose et l'intensité des bandes est comparée à celle du marqueur. La seconde consiste à quantifier les échantillons d'acides nucléiques par mesure d'absorbance à une longueur d'onde de 260 nm. Cette longueur d'onde est choisie car les bases puriques et pyrimidiques absorbent fortement dans l'ultraviolet. La mesure s'effectue dans des microcuvettes en quartz à des dilutions appropriées. Pour chaque mesure, un spectre est tracé entre les longueurs d'onde 250 et 290 nm. La pureté de l'échantillon est évaluée par le rapport des valeurs d'absorbance à 260 et 280 nm. Une valeur de ce rapport comprise entre 1,8 et 2,0 permet d'évaluer une faible contamination en protéines.

2.7. Expression d'*IgTeCe* et *IgEst3*

Pour chaque culture, un suivi de la croissance est réalisé. Lorsque la culture de la souche a atteint une turbidité comprise entre 0,6 et 0,8, il faut procéder à l'induction. Le glucose (2% p/v) est remplacé par du galactose (2% p/v) dans le milieu de culture. Après centrifugation de chaque culture (2 000 g, 10 min, température ambiante), le surnageant contenant le glucose est éliminé. Les cellules sont resuspendues et agitées délicatement dans un milieu SC-U contenant 2 % de galactose. Une seconde centrifugation va permettre l'obtention d'un nouveau culot, qui sera repris dans 100 mL de milieu SC-U contenant du galactose seul. Dès l'introduction du galactose dans le milieu de culture, la culture est considérée comme étant induite, ce point de départ est défini comme le temps zéro d'induction ($T = 0$).

L'expression est réalisée pour *pYE-IgTeCe/pYE-IgEst3* dans un premier volume de 100 mL de culture avec ajout de galactose pour l'induction et dans un second volume de culture de 100 mL avec ajout de glucose (culture non induite). En parallèle de l'expression du plasmide *pYE-IgTeCe/pYE-IgEst3*, deux contrôles ont été réalisés à l'aide d'une construction permettant de produire selon les mêmes conditions la β -galactosidase d'*E.coli*. Le plasmide utilisé est le *pYES2.1-V5-His-LacZ* fourni par Invitrogen™ en complément du plasmide *pYES2.1*. Comme précédemment, une culture est réalisée pour le plasmide *pYES2.1-V5-His-LacZ* avec galactose (contrôle induit) et une culture avec glucose (contrôle non induit).

A partir du $T = 0$, toutes les cultures sont mises sous agitation (170 rpm) à 30 °C durant 24 et 60 heures. Les cellules sont récoltées par centrifugation à 4 000 g pendant 20 min à 4°C. Les masses fraîches des culots de cellules sont alors déterminées. Les culots sont ensuite placés à – 70°C ou utilisés immédiatement.

3. Etude de l'expression chez *Escherichia coli*

3.1. Préambule

L'étude de l'expression dans l'hôte *E. coli* a été entreprise après l'étude de l'expression chez la levure. Dans un souci de clarté et de non redondance, ce paragraphe regroupe des techniques de biologie moléculaire propres à l'étude de l'expression chez *E. coli*.

3.2. Extraction des gènes *IgTeCe* et *IgEst3* depuis le plasmide *pYES2.1* : protocole “ K7-extract ”

L'obtention des gènes *IgTeCe* et *IgEst3* depuis le plasmide *pYES2.1* est réalisée par PCR comme décrit au paragraphe 2.4 en utilisant le couple d'amorces K7EXTRACTFW x K7EXTRACTRV. L'extraction des gènes depuis le plasmide *pYES2.1* est faite afin d'effectuer le clonage ainsi que le séquençage depuis un produit PCR. Les produits PCR sont préparés puis purifiés en utilisant le système Sigma-Aldrich™ comme décrit dans le paragraphe 2.5. Une aliquote est ensuite analysée sur gel pour s'assurer de la purification et de la récupération du produit PCR.

3.3. Clonage et ligation

Les gènes *IgTeCe* et *IgEst3* ont été obtenus à partir du produit PCR “K7extract” en utilisant les amorces présentées dans le tableau 22. Les codons stop présents dans les gènes *IgTeCe* et *IgEst3* ont été supprimés afin d'exprimer les gènes dans des vecteurs d'expression ajoutant une étiquette en région C-terminal. Cette suppression est réalisée à partir du protocole d'amplification décrit dans le paragraphe 2.3.

Dénomination	Sequence 5' -> 3'	Taille en pb	Température d'hybridation °C	Nature de l'amorce
7519-ATG2	ATGGGCGGCCGTCTTCGTCAAAG	23	64,2	Début <i>IgTeCe</i>
7519-STOP2E	GAAAAGAGGCTCCAGGATAAGCTCAGGCG	29	64,2	Modification <i>IgTeCe</i>
6895-ATG2	ATGAGCCCTGGTCATGCTCGCGACACTTC	29	62	Début <i>IgEst3</i>
6895-STOP2E	GTGCACGATCACGGTCACGTACGCG	26	62	Modification <i>IgEst3</i>

Tableau 22: Séquences des amorces utilisées pour caractériser et cloner les gènes *IgTeCe* et *IgEst3*

Sous l'action de la polymérase, les produits PCR générés disposent d'extrémités franches. Les produits de ces amplifications sont directement insérés dans le plasmide *pJET1.2* en suivant les

instructions fournies avec le kit CloneJet® (Fermentas™). La ligation est réalisée dans un volume final de 20 µL. Dans un microtube, sont introduits 10 µL de tampon de réaction (2X), 1 µL d'enzyme de ligation (2000 U/µL) (T4 DNA ligase à haute concentration), 1 µL de produit PCR non purifié (1µg), 1 µL de plasmide *pJET1.2* (50ng) et enfin 7 µL d'eau distillée stérile dépourvue de nucléase. L'ensemble est placé à température ambiante pendant 6 min. Le produit de ligation est ensuite conservé sur glace avant d'être transféré dans les cellules d'*E.coli* chimiocompétentes.

3.4.Préparation de cellules d'*Escherichia coli* et transformation avec le plasmide de clonage.

Préparation de cellules chimiocompétentes TOP 10

Afin de rendre les bactéries compétentes, c'est-à-dire capables d'accepter un ADN exogène, celles-ci sont incubées à 37 °C sur une boîte de Petri contenant du milieu LB gélosé. A partir d'une colonie, un ensemencement est effectué dans un volume de 10 mL de milieu LB, et placé sous agitation à 37°C pendant 15 heures. Le lendemain, une dilution de la préculture au 1/100^e est transférée dans 200 mL de milieu LB et remis en culture à 37°C sous agitation. Lorsque la turbidimétrie à 595 nm atteint environ 0,5, la culture est transférée dans des tubes coniques de 50 mL. Les tubes sont maintenus sur la glace pendant 15 min. Après centrifugation à 4 000 g durant 10 min à 4°C, les bactéries sont reprises dans un volume de MgCl₂ (100 mM). Les cellules sont délicatement resuspendues et préservées dans la glace durant 30 min. Après centrifugation dans les mêmes conditions que celles précédemment décrites, le culot cellulaire est resuspendu délicatement dans 1/20^e du volume initial avec une solution de CaCl₂ (85 mM) / glycérol (15 %) et conservé dans la glace pendant 1 heure. Enfin, en chambre froide, les suspensions sont réparties dans des cryotubes par volume de 100 µL et stockées à – 70°C.

Transformation par choc thermique

En conditions stériles, 100 µL de bactéries chimiocompétentes TOP 10, préalablement décongelés dans la glace, sont ajoutés à 5 µL de produit de ligation dans un microtube de 1,5 mL. L'ensemble est placé sur la glace pendant 30 min. Le transfert du plasmide dans la bactérie est provoqué par un choc thermique de 45 secondes à 42°C. Le microtube est de nouveau transféré dans la glace et 400 µL de milieu enrichi SOC (tryptone 2%, extraits de levure 0,5 %, NaCl 8,5 mM, KCl 2,5 mM, MgCl₂ 10 mM et glucose 20 mM) sont ajoutés avant étalement sur boîtes de Petri.

L'étalement s'effectue au moyen de billes de verre stériles sur boîtes LB-agar complémentées en antibiotiques nécessaires à la sélection. Deux volumes de 50 et 150 μ L de suspension sont ensemencés. Les boîtes de Petri sont ensuite incubées 24 heures à 37°C.

3.5.Extraction de plasmides à partir d'*Escherichia coli*

3.5.1. Mini préparation de plasmides pour la sélection de clones positifs

Un nombre suffisant de colonies présentes sur boîtes de Petri à l'issue du temps d'incubation est prélevé avec la pointe d'un cône stérile de micropipette. Celui-ci est éjecté dans un tube conique de 15 mL contenant 4 mL de LB + antibiotique puis placé sous agitation à 37°C pendant 15 heures. Le lendemain, l'ensemble des tubes où un développement s'est produit va être analysé par la technique de mini-préparation suivant le protocole "TENS" (Zhou *et al.*, 1990). Ce protocole suit une technique de lyse alcaline et permet un isolement rapide des plasmides bactériens. Un prélèvement de 1,5 mL de suspension bactérienne est déposé dans un microtube, le volume de suspension restant est placé à 4°C et servira de préculture pour le repiquage des clones positifs. Le microtube est centrifugé à 4000 g à température ambiante pendant 10 min. Le surnageant est retiré sans assécher le culot. Le culot est repris dans 300 μ L de tampon TENS (1 M Tris pH 8 (1 %), 20 % SDS (2,5 %), 10 N NaOH (1 %), H₂O) puis resuspendu par retournement durant 5 min avant d'ajouter 100 μ L d'acétate de sodium (3 M pH 5,2) et d'homogénéiser à nouveau. Une nouvelle centrifugation est réalisée dans des conditions identiques, le surnageant est transféré dans un nouveau tube. La précipitation est réalisée par l'ajout d'un volume d'éthanol absolu filtré (à l'aide d'un filtre seringue de 0,22 μ m) et froid. Le culot d'ADN, récupéré par centrifugation, est lavé délicatement avec de l'éthanol 70% et centrifugé une nouvelle fois. L'éthanol 70 % est prélevé délicatement par aspiration à l'aide d'une pompe à vide afin d'obtenir un culot asséché. L'obtention d'une mini-préparation d'ADN plasmidique est complète en reprenant ce culot dans 30 μ L d'H₂O stérile. La qualité de la préparation d'ADN plasmidique est déterminée par visualisation sur un gel d'agarose (dépôt de 2 μ L d'échantillon pour un volume total de dépôt de 10 μ L).

3.5.2. Midi préparation de plasmides pour sous-clonage et séquençage

Les plasmides sont préparés à partir de plus grandes quantités de cellules. L'ADN plasmidique est extrait à partir d'une culture bactérienne réalisée dans un volume de 50 mL de milieu LB + antibiotique. Cette extraction est faite à l'aide du kit Wizard Plus Midiprep DNA purification system (Promega®) en suivant les recommandations du fournisseur. L'élution de l'ADN est répétée deux fois

et non regroupée. La qualité et la quantité de l'ADN sont également déterminées. Ce protocole est aussi utilisé pour l'obtention d'un échantillon à envoyer sur une plateforme de séquençage.

3.6. Validation d'un clone

La validation de clones permet de vérifier la qualité du clonage et de s'assurer de la présence d'un fragment inséré au sein d'un plasmide. Cette validation comporte la vérification de plusieurs paramètres par le biais de différentes techniques, la taille de l'insert par digestion enzymatique, son identité et enfin son intégrité par séquençage.

3.6.1. Validation par digestion enzymatique

La digestion enzymatique est la technique de base pour la sélection de clones pouvant être positifs. La digestion est effectuée sur des échantillons d'ADN plasmidique. Le plasmide *pJET1.2* est digéré par l'enzyme *HindIII* dans un volume total de 20 µL. Dans un microtube de 1,5 mL, sont introduits 50 à 500 ng d'ADN plasmidique, 1 µL d'enzyme de restriction *HindIII* (Sigma-Aldrich™) (10U/µL), 2 µL de tampon de réaction 10X SB fourni par Sigma-Aldrich™ (10 mM Tris-HCl, 100 mM NaCl, 5 mM MgCl₂, pH 8) et de l'eau ultrapure stérile en quantité suffisante pour 20 µL. Le milieu réactionnel est incubé pendant 1 heure à 37 °C puis inactivé pendant 20 min à 65 °C. La digestion est contrôlée sur gel d'agarose par séparation électrophorétique. La taille attendue pour le produit de digestion linéarisé est de 4 217 bp, correspondant au plasmide *pJET1.2* (2974 bp) contenant l'insert *IgTeCe* (1243 bp).

3.6.2. Validation par séquençage

Le séquençage est réalisé en utilisant les amorces du tableau 23. Le résultat du séquençage va permettre de vérifier l'orientation de la séquence dans le vecteur *pJET1.2*.

Dénomination	Sequence 5' -> 3'	Taille en pb	Température d'hybridation °C	Nature de l'amorce
7519-L2	CTGCTGCCAGGGATTAAATATGTCTT	26	57,3	Interne à <i>IgTeCe</i>
7519-ATG	ATGGGCGGCCGTCTTCGTCAAAGCCC	26	68,9	Début <i>IgTeCe</i>
7519-STOP	CTAGAAAAGAGGCTCCAGGATAAGCTCAGG	30	61,4	Fin <i>IgTeCe</i>
pJET1.2 Forward	CGACTCACTATAGGGAGAGCGGC	23	60,3	Séquençage
pJET1.2 Reverse	AAGAACATCGATTTTCCATGGCAG	24	56,3	Séquençage
T7-Promoter	TAATACGACTCACTATAGGG	20	47	Séquençage

Tableau 23: Amorces pour le séquençage dans le plasmide *pJET1.2*

3.7. Clonage dans les vecteurs d'expression

3.7.1. Amplification et ajout de sites de restriction

Deux méthodes sont mises en place pour le clonage du gène *IgTeCe*. La première méthode a consisté en l'ajout de sites de restriction aux extrémités d'*IgTeCe*. Pour réaliser ces ajouts, le gène *IgTeCe* a été amplifié par le couple d'amorces 7519FWNCOI et 7519RVNOTI et par le couple 7519PTYFWDNDEI et 7519RVNOTI (Tableau 24). Ces amplifications sont réalisées avec une température d'hybridation de 60 °C, un nombre de cycle égal à 35 et un temps d'élongation fixé à 70 secondes. Deux constructions ont donc été prévues : une avec les extrémités *NcoI* en 5' et *NotI* en 3' et l'autre avec les extrémités *NdeI* en 5' et *NotI* en 3'. Ces amplicons pourront être insérés dans les vecteurs d'expression *pET28a+* et *pET52b+*.

Dénomination	Sequence 5' -> 3'	Taille en pb	Température d'hybridation °C	Nature de l'amorce
7519FWNCOI	CGA CC ATGGGCGGCCGTCTTCGTC	24	67,8	Ajout site de restriction
7519PTYFWDNDEI	CGA CAT ATGGGCGGCCGTCTTCGTC	25	65,5	Ajout site de restriction
7519RVNOTI	CGAGCGGCCCGCGAAAAGAGGCTCCAGGA	28	71,1	Ajout site de restriction

Tableau 24: Amorces pour l'ajout de sites de restriction à *IgTeCe*

La seconde méthode va utiliser les sites de restriction disponibles sur le plasmide de clonage *pJET1.2*. Des bouts cohésifs formés de part et d'autre du gène *IgTeCe* permettront ainsi le clonage dans les vecteurs d'expression (Geszvain and Tebo, 2010).

3.7.2. Digestion enzymatique pour l'insertion dans un plasmide d'expression

Les enzymes de restriction utilisées (*NcoI*, *NotI*, *NdeI*) ont été fournies par Fermentas® et par Sigma-Aldrich® (10 U/μL). La digestion des plasmides a été réalisée par adaptation des protocoles proposés par les fournisseurs. Le tableau 25 précise les sites de restriction pour le sous-clonage dans les différents vecteurs d'expression.

Plasmide d'expression	Site 5' de clonage	Site 3' de clonage	Etiquette ajoutée
<i>pET28a+</i>	<i>NcoI</i>	<i>NotI</i>	His ₆ en C-terminal
<i>pET28a+</i>	<i>NdeI</i>	<i>NotI</i>	His ₆ en C-terminal et N-terminal
<i>pET52b+</i>	<i>NcoI</i>	<i>NotI</i>	His ₁₀ +site de clivage à la thrombine en C-terminal

Tableau 25: Sites de restriction pour le sous-clonage en plasmide d'expression

La digestion des plasmides est effectuée dans un volume final de 50 μL en présence de 15 μL de vecteurs à digérer, 10% v/v d'enzyme de restriction, 30% v/v de tampon de digestion. L'ensemble est complété par de l'eau distillée stérile. La digestion des produits PCR est effectuée en ajoutant 6,5% v/v d'enzyme de restriction et 6,5% v/v de tampon de digestion à 10 μL de produits PCR. De l'eau

distillée stérile est ajoutée pour obtenir un volume final de 30 μL . Chaque milieu réactionnel est incubé à 37°C pendant 90 min. Les milieux réactionnels sont par la suite inactivés à 80°C pendant 25 min. Une aliquote de 10 μL est déposée sur gel d'agarose afin de vérifier la bonne digestion des fragments, si tel est le cas une extraction du fragment est réalisée.

3.7.3. Ligation et transformation

La ligase T4 de NEB® a été privilégiée pour la réalisation de la ligation des inserts dans les plasmides d'expression. Lors de cette étape, les recommandations du fournisseur ont été utilisées. La réaction a été réalisée dans un microtube de 1,5 mL avec un volume final de 20 μL . Le tampon fourni avec l'enzyme permet l'apport d'ATP (1mM final), molécule essentielle à la réaction. La ligation s'effectue en présence du tampon en concentration finale 1X, de l'insert, du vecteur et 1 μL de ligase (400 U/ μL). Deux proportions entre l'insert et le vecteur ont été testées, l'une avec la proportion 4/1 et l'autre 2/1. Le tube est par la suite incubé à température ambiante pendant 1 heure. La validation des clones dans les souches de clonage puis les souches d'expression suit les protocoles décrits dans les paragraphes 3.4, 3.5.1, 3.5.2 et 3.6.

3.8. Validation de l'insertion dans les plasmides d'expression

3.8.1. Validation par séquençage

Le séquençage est réalisé comme au paragraphe 3.6.2. Les travaux de séquençage ont utilisé les amorces du tableau 26.

Dénomination	Sequence 5' -> 3'	Taille en pb	Température d'hybridation °C	Nature de l'amorce
7519-L2	CTGCTGCCAGGGATTAAATATGTCTT	26	57,3	Interne à <i>IgTeCe</i>
7519-ATG	ATGGGCGGCCGTCTTCGTCAAAGCCC	26	68,9	Début <i>IgTeCe</i>
7519-STOP	CTAGAAAAGAGGCTCCAGGATAAGCTCAGG	30	61,4	Fin <i>IgTeCe</i>
T7-Promoter	TAATACGACTCACTATAGGG	20	47	Séquençage pET28/pET52
T7-Term	CTAGTTATTGCTCAGCGG	18	49,7	Séquençage pET28/pET52

Tableau 26: Amorces pour le séquençage dans les plasmides d'expression

3.9. Préparation de cellules d'*Escherichia coli* BL21 et transformation avec les plasmides d'expression

3.9.1. Préparation de cellules électrocompétentes

A partir d'une préculture en milieu LB, 5 mL sont transférés dans un volume de culture de 500 mL de milieu LB. La culture est placée sous agitation à 37°C jusqu'à atteindre une turbidimétrie à 595 nm comprise entre 0,5 et 0,7. La suspension cellulaire est alors placée dans la glace pendant 20 min. Après centrifugation à 4 000 g durant 15 min à une température de 4°C, le culot est resuspendu dans

500 mL d'une solution stérile d'eau déminéralisée à 10 % de glycérol (v/v) et de nouveau centrifugé dans les mêmes conditions. Le culot obtenu est resuspendu dans 250 mL de la même solution à 10 % de glycérol (v/v) et une nouvelle centrifugation est réalisée. Le culot est resuspendu dans 20 mL de solution à 10 % de glycérol (v/v). Après centrifugation, le culot obtenu est resuspendu dans 1 à 2 mL de solution à 10 % de glycérol (v/v), les suspensions sont réparties dans des cryotubes par volume de 100 μ L et stockées à -70°C .

Transformation par choc électrique

En condition stérile, 100 μ L de bactéries électrocompétentes sont ajoutés à 5 μ L de produit de ligation. L'ensemble est placé dans une micro-cuve d'électroporation de 2 mm (préalablement refroidie dans la glace). La cuve d'électroporation est ensuite placée dans l'électroporateur (Electroporator 2510 de la société Eppendorf). Les conditions d'électroporation sont fixées à 23 kV/cm. L'impulsion électrique est immédiatement suivie de l'ajout de 400 μ L de milieu SOC. Une fois que l'ensemble des électroporations a été réalisé, les suspensions obtenues sont de nouveau placées dans des microtubes de 1,5 mL avant d'être mis sous agitation à 37°C pendant une heure. La suspension obtenue est directementensemencée par étalement sur boîte de Petri.

3.9.2. La culture d'*Escherichia coli* – tests d'expression

L'expression est étudiée dans différentes conditions. Chaque clone d'expression (association d'un plasmide d'expression contenant la séquence *IgTeCe* et d'une souche d'expression) dispose d'un système d'induction à l'IPTG. L'induction, à partir d'une solution mère à 1 M d'IPTG, est réalisée directement dans le milieu de culture lorsque la densité cellulaire atteint une turbidimétrie comprise entre 0,6 et 0,8 à 595 nm. Le tableau 18 décrit l'ensemble des conditions d'expression testées pour l'expression du gène *IgTeCe* chez *E.coli*. Pour chaque clone d'expression, la préculture est réalisée à une température de 37°C sous agitation alors que les cultures se feront à plusieurs températures. Chaque culture induite et non induite est réalisée en triplicat. Les différents paramètres testés pour les cultures induites sont indiqués dans le tableau 27. Les cultures non induites sont traitées dans les mêmes conditions sans IPTG.

Clone	Température durant induction	Temps d'induction	Concentration finale IPTG	Choc induction
<i>pET28 Rosetta™ 2</i>	37°C	3; 6; 8 heures	0,4 mM; 1 mM	Aucun
	30°C	8 heures	1 mM	Bain eau tempérée
	16°C	16 heures	1 mM	Mélange eau-glacé
<i>pET28 Star™ 2</i>	37°C	3; 6 heures	1 mM	Aucun
<i>pET52 Rosetta™ 2</i>	37°C	3; 6 heures	0,4 mM; 1 mM	Aucun
	30°C	8 heures	1 mM	Bain eau tempérée
	16°C	16 heures	1 mM	Mélange eau-glacé

Tableau 27: Conditions d'expression testées

Dans le cas d'une induction à 30 °C, les Erlenmeyers sont placés une fois la turbidité voulue atteinte dans un bain d'eau tempérée à 30 °C sans agitation pendant 30 min avant l'ajout d'IPTG. De la même manière pour la culture induite à 16 °C, les Erlenmeyers sont placés pendant 30 min dans un mélange eau/glacé pour diminuer drastiquement la température du milieu avant induction.

A l'issue des temps d'induction, la densité cellulaire est estimée et la biomasse est récupérée par centrifugation à 4 000 g pendant 20 min à 4°C. Les culots de cellules sont alors pesés avant d'être placés à – 70°C ou utilisés immédiatement.

4. Recherche de transcrits (RT-PCR)

4.1. Préparation d'ARN totaux

La préparation d'ARN total, dans le but d'effectuer une étude RT-PCR, a été entreprise pour vérifier la présence du transcrit *IgTeCe* ou *IgEst3* lors d'une culture induite et son absence pour une culture non induite. Les ARN totaux ont été extraits à la fois pour les clones de levure et pour les clones de bactéries grâce au kit SV-total RNA de Promega™ en suivant les instructions du fournisseur indiquées dans le tableau 28. Les modifications du protocole ont été apportées pour la lyse de chaque type cellulaire en suivant les informations complémentaires fournies et en utilisant pour la levure la lyticase (50 U) et pour la bactérie le lysozyme (0,5 mg/mL).

ARN total depuis la levure <i>S. cerevisiae</i>	ARN total depuis la bactérie <i>E. coli</i>
50 mg de cellules sont repris dans 100 µL de tampon SEB ("Sorbitol EDTA Buffer" ; 1M sorbitol, 0,1M EDTA (pH 7,4)), et incubés avec de la lyticase (50 U final) pendant 30 min à 30 °C.	Les 50 mg de cellules sont repris dans 100 µL de TE (Tris-EDTA) pH 8,0 contenant 0,5mg/mL de Lysozyme et incubés à température ambiante durant 10 min
Ensuite, 75 µL de tampon de lyse (kit promega) et 350 µL de tampon de dilution d'ARN sont ajoutés et une agitation douce est effectuée durant 5 min.	
La suspension est alors centrifugée à vitesse maximum et à température ambiante, le surnageant est transféré dans un nouveau tube contenant 200 µL d'éthanol absolu	Une suspension éclaircie est obtenue à l'issue de l'agitation et est transférée dans un nouveau tube contenant 200 µL d'éthanol absolu

Tableau 28: Usages spécifiques du kit d'extraction d'ARN total Promega

Le mélange est alors transféré sur l'assemblage de colonne fourni par le kit Promega. Après une série de centrifugations et de lavages, les échantillons sont traités à la DNase I dans les conditions du kit. Enfin, l'ARN total est récupéré par ajout de 100 µL d'eau dépourvue de DNase ou de RNase et placé à -20°C.

4.2. Obtention de la matrice ADN complémentaire

L'ADNc a été obtenu grâce au kit RT-PCR Phusion® de la société Fermentas™. La transcription inverse, réaction qui consiste à obtenir une molécule d'ADN depuis une matrice constituée d'ARN a été réalisée par la "reverse transcriptase" M-MuLV Reverse Transcriptase (Fermentas™). Les fragments d'ADN générés sont appelés ADN complémentaires (ADNc) et ont été à leur tour utilisés comme matrice afin d'effectuer une amplification spécifique d'un ADNc précis. Dans un microtube PCR, 4 µL d'ARN total (1 µg) sont introduits, puis 1 µL d'un mélange de dNTP (0,5 mM final) et 1 µL d'amorces oligo-dT (100 ng final). L'ensemble est complété par 4 µL d'eau ultrapure traitée contre les nucléases. Le mélange est incubé pendant 5 min à 65 °C dans le thermomixeur de manière à prédénaturer l'ARN. Deux microlitres de tampon de rétrotranscription (10X), 2 µL d'un mélange enzymatique contenant la transcriptase inverse et 6 µL d'eau distillée stérile sont introduits dans le tube. Celui-ci est placé de nouveau dans le thermomixeur pendant 10 min à 25 °C (extension d'amorce), 30 min à 40 °C (synthèse d'ADNc), 5 min à 85 °C pour terminer la réaction. Enfin, le mélange est ramené à 4°C pendant 20 min.

4.3. Evaluation de la RT-PCR

La réaction de RT-PCR génère des fragments amplifiés qui sont analysés par migration électrophorétique en gel d'agarose (1,5 % (p/v)) dans un tampon Tris Acetate EDTA 1X (TAE). La taille des fragments est déterminée à l'aide du marqueur de taille GeneRuler 1 kb DNA Ladder de la société Fermentas™. Les échantillons sont chargés (5/6^e du volume final) avec le tampon 6X DNA

Loading Dye (1/6e du volume final) de Fermentas™. La migration est réalisée sous une tension constante de 100 V. Afin de visualiser les fragments d'ADN aux rayons UltraViolet (UV), le gel d'agarose est trempé dans un bain contenant du bromure d'éthidium. La prise de vue est réalisée sous UV sur un appareil GelDoc™ de BioRad®. La taille attendue du produit RT-PCR est validée en la comparant à celle obtenue à partir du plasmide d'expression contenant le gène. Les amorces testées sont présentées dans le tableau 29.

Dénomination	Sequence 5' -> 3'	Taille en pb	Température d'hybridation °C	Nature de l'amorce
6895-ATG2	ATGAGCCCTGGTCATGCTCGCGACACTTC	29	62	Début <i>IgEst3</i>
6895-STOP2E	GTGCACGATCACGGTCACGTCACGCG	26	62	Fin <i>IgEst3</i>
7519-ATG2	ATGGGCGGCCGTCTTCGTCAAAG	23	64	Début <i>IgTeCe</i>
7519-STOP2E	GAAAAGAGGCTCCAGGATAAGCTCAGGCG	29	64,2	Fin <i>IgTeCe</i>

Tableau 29: Amorces PCR pour la RT-PCR

5. Techniques de biochimie

5.1.Extraction des protéines de *Saccharomyces cerevisiae*

Les protéines ont été extraites depuis la levure par sonication avec l'appareil Hielscher UP50H. Chaque culot est pesé et pour chaque gramme de cellule un volume de 4 mL de tampon de lyse est utilisé. Différents tampons de lyse ont été testés et deux séries de cycles de sonication sont nécessaires pour extraire les protéines (Tableau 30). La première série appelée “présonication” est composée de 5 cycles de sonication de 45 secondes, chaque cycle étant précédé d'une pause de 60 secondes sur glace. La seconde série de cycles permet, après fragilisation de la cellule de levure, d'obtenir un lysat éclairci. Elle est identique à la première. La lyse des cellules de levure est vérifiée au microscope optique (comptage en cellule de Malassez). Les lysats obtenus sont centrifugés à 25 000 g pendant 45 min à 4°C. Le culot est préservé à -70°C et les surnageants en aliquotes à -20°C.

Tampon de lyse	Composition	
Tampon 1	100 mM Tris HCl, 50mM NaCl	5mM 2-mercaptoéthanol & 3 mM benzamidine, pH 8
Tampon 2	100 mM Tris HCl, 100 mM KCl	
Tampon 3	20 mM HEPES, 100 mM KCl	
Tampon 4	20 mM HEPES, 50 mM NaCl	
Tampon 5	20 à 100 mM phosphate de sodium	

Tableau 30: Tampons de lyse utilisés pour l'extraction de protéines de levure

5.2.Extraction des protéines d'*Escherichia coli*

L'extraction des protéines depuis les culots cellulaires de bactéries a également été réalisée par sonication. La préparation a été préalablement refroidie dans la glace pendant 30 min avant

sonication. La sonication a été faite avec l'appareil Hielscher® UIS 250. Chaque culot cellulaire a été resuspendu dans un volume de tampon proportionnel à la masse fraîche de cellules (6 mL/g). Les suspensions de cellules en tampon de lyse (Tableau 31) sont soniquées par cycle de 30 secondes d'impulsion puis 45 secondes de repos sur glace jusqu'à éclaircissement des suspensions cellulaires. Un double contrôle a été réalisé afin de confirmer le bon cassage des cellules (coloration de Gram et mesure de turbidité). Une fois le lysat obtenu, celui-ci est centrifugé pendant 45 min à 25 000g (4°C). Les surnageants sont préservés en aliquotes à -20°C tandis que le culot de débris est traité par extraction en tampon dénaturant.

Tampon de lyse	Composition
Tampon 1	100 mM Tris HCl, 500mM KCl, 5mM 2-mercaptoéthanol & 1 mM benzamidine, pH 8
Tampon 2	50 mM phosphate sodium, 300 mM NaCl, pH 8
Tampon 3	PBS 1x (Sambrook <i>et al.</i> 2001) pH 8
Tampon 4	TrisHCl 50mM, pH 8

Tableau 31: Différents tampons de lyse utilisés sur les cellules d'*E. coli*

Extraction en tampon dénaturant

Les tampons dénaturants testés sont issus des différents tampons de lyse décrits dans le tableau 23 dans lesquels 8 M d'urée sont ajoutés. Après lyse en tampon non dénaturant, le culot réservé à -70°C a été resuspendu dans une solution (750 µL/g de culot cellulaire au départ) de tampon dénaturant. L'ensemble a été placé sous agitation à température ambiante (25°C) pendant 60 min (750 rpm). Les microtubes contenant les échantillons sont centrifugés durant 25 min à 25 000 g. Le surnageant est préservé pour analyse en gel de polyacrylamide (en condition dénaturante) ou en Dot Blot de la même manière que pour les échantillons obtenus en condition non dénaturante.

5.3. Dosage de protéines

La quantité de protéines totales pour chacun des échantillons obtenus a été dosée par la méthode de Bradford (Bradford, 1976). Le réactif de Bradford provient de la société Bio-Rad. Les mesures ont été faites dans des micro-cuves de 1 mL. Une courbe standard est obtenue par utilisation de sérum d'albumine bovine (SAB). Les concentrations de SAB sont comprises entre 0 et 25 µg/mL.

5.4. Séparation électrophorétique

5.4.1. SDS-PAGE

L'électrophorèse en gel de polyacrylamide est réalisée selon Sambrook et Green (Tableau 32) (Green *et al.*, 2012). La solution de polyacrylamide a été fournie par la société Roth. L'électrophorèse a lieu

en condition dénaturante, en présence de sodium dodécylsulfate (SDS) qui donne aux protéines une charge globale négative et permet de les séparer selon leur taille. Les gels sont préparés entre deux plaques de verre (épaisseur de gel 0,75 mm) à l'aide d'un appareil Bio-Rad® Protean 3. Les cuves d'électrophorèse sont également fournies par BioRad. L'électrophorèse est réalisée dans un tampon de migration 1X préparé à partir d'une solution mère 10X contenant 144 g de glycine, 30 g de tris-base et 10 g de SDS par litre d'eau distillée (TGS 1X, pH 8,3). Un volume compris entre 10 et 25 µL selon l'échantillon est déposé dans les puits. Les protéines sont dénaturées dans le tampon décrit par Laemmli (Tris-HCl 100 mM pH 8,5 mM 2-mercaptoéthanol, 10 % Glycérol, 2 % SDS) et placées pendant 8 min à 98°C. La migration est réalisée à température ambiante dans un champ électrique (200 V) en présence de tampon de migration pour une durée de 50 min (Laemmli, 1970).

Gel de séparation 12 %	Gel de concentration 4,25 %	Tampon de migration
Solution d'Acrylamide/BisAcrylamide (37,5/1) 30% : 8,0 mL	Solution d'Acrylamide/BisAcrylamide (37,5/1) 30% : 750 µL	Tris : 25 mM
Tris HCl pH 8,8 1.5 M : 5,0 mL	TrisHCl pH 6,8 0.5 M : 750 µL	Glycine : 192 mM
H ₂ O : 6,6 mL	H ₂ O : 3,6 mL	
Sodium Dodécyl Sulfate (10%) : 200 µL	SDS (10%) : 50 µL	
Peroxydisulfate d'ammonium (APS) (10%) : 200 µL	APS (10%) : 50 µL	SDS : 0,1%
Tetramethylethylenediamine (Temed) : 40 µL	Temed : 25 µL	

Tableau 32: Composition des gels et tampons nécessaires à l'électrophorèse de protéines

Un marqueur de poids moléculaire est utilisé à chaque électrophorèse de protéines. Si le gel n'est pas transféré sur membrane, un marqueur non coloré sera utilisé. Dans le cas contraire, un marqueur pré-coloré permettra de vérifier le transfert. Les références et visualisations des différents marqueurs de poids moléculaire sont présentées dans le tableau 33.

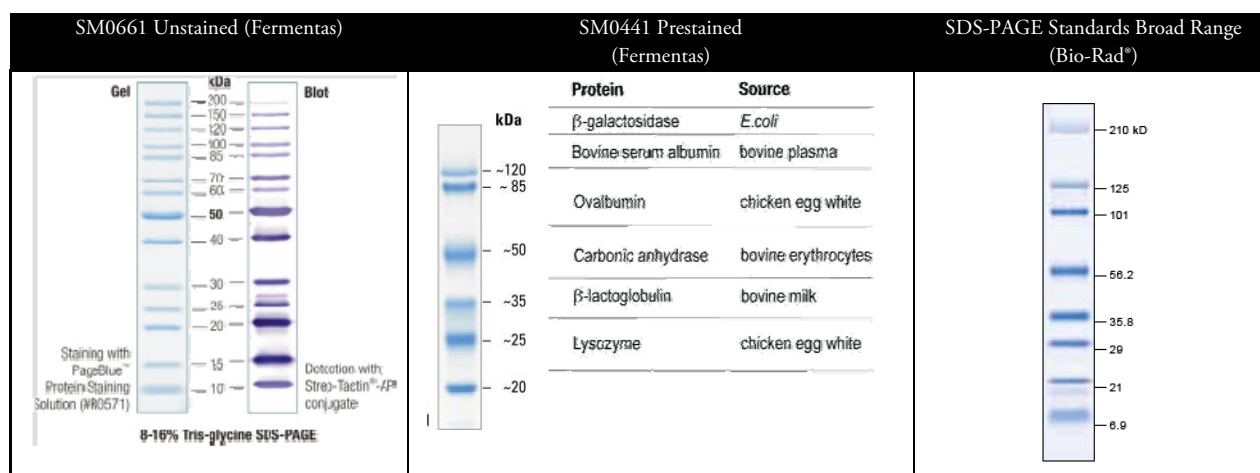


Tableau 33: Marqueurs de tailles utilisés pour les électrophorèses de protéines

5.4.2. Coloration des protéines

Toutes les phases d'incubation sont effectuées sous agitation. La coloration des protéines sur SDS-PAGE est réalisée immédiatement après migration du gel par incubation dans un bain de fixation (20 mL/mini gel) composé de 2 % d'acide phosphorique et 50 % d'éthanol pendant une nuit. Le gel est ensuite incubé 2 fois 30 min dans un bain de rinçage (même volume). Le bain de rinçage est composé de 2 % d'acide phosphorique. Le gel est ensuite placé dans un bain de pré-trempe composé de 2 % d'acide phosphorique, 17 % d'éthanol et 15 % de sulfate d'ammonium. Après 15 min d'agitation, 150 µL de solution de coloration contenant 0,05 % de Bleu de coomassie R 250 (préparé en éthanol absolu) sont ajoutés. Le gel est incubé jusqu'à apparition de bandes protéiques. Une fois que les bandes sont correctement colorées, le gel peut être décoloré dans un bain d'eau distillée en présence d'un papier absorbant afin de retenir l'excédent de colorant.

5.5. Détection immunologique

5.5.1. Dot blot

Le Dot blot permet de détecter, et d'identifier les échantillons protéiques contenant la protéine d'intérêt. La protéine d'intérêt sera repérée par hybridation spécifique d'un anticorps donné. Une peroxydase est présente sur l'anticorps et une réaction chromogène permet alors de visualiser la présence ou non de la protéine recherchée. L'analyse s'effectue dès l'obtention d'extrait protéique ou à partir d'extraits conservés à -20°C. Une membrane de nitrocellulose est utilisée de manière à immobiliser l'extrait protéique. Cette immobilisation est réalisée par dépôt et séchage (à température ambiante) de 4 µL d'extrait protéique. Le dépôt devra s'effectuer délicatement et lentement de manière à minimiser l'étendue du dépôt et le bruit de fond. Après séchage de la membrane, 10 mL d'une solution de blocage composée de 5 % SAB dans du PBS-Tween (0,5%) sont ajoutés. La saturation s'effectue sous agitation durant une heure à température ambiante ou une nuit à une température comprise entre 4 et 8 °C. A l'issue du temps de saturation, la solution de blocage est retirée et la membrane est incubée trois fois 10 min dans le même volume de solution de PBS-Tween seul. La membrane est mise sous agitation pendant 4 à 5 heures avec l'anticorps primaire couplé à la peroxydase de raifort (HRP) (1/5000^e) dans la solution de blocage. L'hybridation est révélée par action de l'HRP sur le réactif (DiaminoBenzidine) DAB du kit Invitrogen™. Après trois lavages de 10 min dans une solution de PBS-Tween puis deux fois 10 min dans du PBS seul, la membrane est séchée 30 secondes à l'air. Le kit DAB Invitrogen™ permet de révéler la présence de protéines détectées par l'anticorps Anti-His.

5.5.2. Western blot

Le Western blot vient compléter l'électrophorèse sur SDS-PAGE. A la suite de la migration, le haut du gel est enlevé. La partie séparative du gel de polyacrylamide est alors déposée en contact de la membrane de nitrocellulose (Biorad®) préalablement imbibée de tampon de transfert (Tris Glycine SDS 1X, 20% éthanol). L'immobilisation des protéines séparées par électrophorèse sur la membrane est réalisée par électrotransfert. Les protéines chargées négativement se transfèrent depuis le gel vers la membrane placée du côté de la cathode. Cette étape utilise un système de transfert MiniTransBlot (Biorad®) à une tension de 100 V pendant 1 heure. Le transfert a lieu en présence d'un bloc réfrigérant et d'une agitation magnétique pour éviter toutes interférences ioniques et de température comme recommandé par le fournisseur. A l'issue du transfert, la membrane est traitée de la même manière qu'à la suite du dépôt d'échantillon et du séchage décrits pour le Dot blot. Une fois la bande révélée, la membrane est à nouveau placée dans du tampon PBS pour stockage et prise de vue.

5.6. Chromatographie et purification

5.6.1. Préambule

La purification des protéines a été réalisée par chromatographie d'affinité. Le gène *IgTeCe* est en effet cloné dans deux vecteurs d'expression permettant la production de la protéine *IgTeCe* avec une étiquette poly-histidine. Le gène *IgEst3* est également cloné dans un vecteur d'expression permettant la production de la protéine *IgEst3* avec une étiquette poly-histidine en C-terminal. Dans l'ensemble des constructions, cette étiquette est composée de 6 ou 10 histidines et a une forte affinité pour les ions divalents comme le cobalt ou le nickel.

Deux systèmes de chromatographie ont été testés. Le premier utilise la force centrifuge par l'intermédiaire de microcolonnes centrifugeables "Spin Column" fournies par la société Sigma-Aldrich™ et le second utilise la pression atmosphérique dans des colonnes en polypropylène "Midi-prep" de la société Bio-Rad™. Selon le système choisi, les volumes de résine testés varient entre 50 et 2 000 µL. Deux résines ont été testées. La première a été conçue à partir de l'ion Nickel de référence HIS-Select®Nickel Affinity Gel de Sigma® appelée « NiNTA » et la seconde à partir de l'ion Cobalt de référence HiTrap™ TALON® Superflow™ de GE Healthcare™ appelée « TALON ».

5.6.2. Tampons utilisés

La composition des différents tampons utilisés reprend celles des tampons de lyse. Tous ces tampons doivent être conservés à 4°C et maintenus sur la glace pendant le procédé de purification. La purification par affinité a fait l'objet de nombreuses conditions expérimentales (Tableau 34).

Tampon 1		Tampon 2	Tampon 3	Tampon 4	Informations complémentaires
Nature du tampon	Tampon de lyse 1	Tampon recommandé avec les résines de purification	Tampon de lyse 2	Tampon recommandé pour tester l'activité	
Composition du tampon de base	100 mM Tris HCl, 500mM KCl, 5mM 2-mercaptoéthanol & 1 mM benzamidine, pH 8	50 mM phosphate sodium, 300 mM NaCl. pH 8	PBS 1x (Sambrook <i>et al.</i> 2001) pH 8	TrisHCl 50mM, pH 8	
Equilibration / Lavage	0 mM				
Elution 1	20 mM				
Elution 2	50 mM				
Elution 3	250 mM				
Nettoyage colonne	1 M				
Dénaturation	Tampon de Laemmli				

Tableau 34: Tampons utilisés pour la chromatographie d'affinité

5.6.3. Procédure de purification

Les échantillons de protéines purifiées ont été préparés en suivant et adaptant les protocoles fournis par les fabricants Sigma-Aldrich™ et GE Healthcare™.

Les volumes à utiliser au cours de la chromatographie ont été établis sur la base du volume de résine utilisé (Column Volume : CV). Les volumes de tampons utilisés à chaque étape de la chromatographie seront exprimés en nombre de CV. Dans un premier temps, le volume CV de résine est placé dans le système de purification et 10 CV du tampon de base sont utilisés pour rincer la résine et ainsi retirer la solution de stockage. Dans un second temps, la résine est équilibrée par 5 CV de tampon d'équilibration. Une fois les volumes de tampon éliminés, la résine est resuspendue dans 1 CV de tampon d'équilibration et est incubée avec 1 CV (100 µg à 1mg) d'extrait protéique dans un tube. L'ensemble est placé à température ambiante sous agitation pendant une heure. Cette étape a permis de maximiser l'interaction entre la résine et les protéines susceptibles de s'y attacher. A l'issue de l'agitation, le mélange est remplacé dans la colonne et la fraction non retenue est récupérée. La résine est par la suite lavée par 10 CV de tampon d'équilibration et l'ensemble des fractions est récupéré et analysé en tant que "lavage". L'élution des protéines est réalisée à différentes concentrations d'Imidazole (Tableau 34), chaque volume d'élution est récupéré. Pour chaque essai

de chromatographie, une aliquote de chaque fraction (non retenu, lavages, éluions) est analysé par SDS-PAGE et Western blot.

6. Caractérisation des enzymes

6.1. Concentration et dialyse

Concentration

Les échantillons protéiques ont été concentrés sur membrane par centrifugation. Les membranes (Vivaspin 500 et Vivaspin 15 de la société Sartorius™) ont permis une séparation des molécules de poids moléculaire inférieur et supérieur à 30 kDa. Chaque échantillon est déposé en suivant la procédure recommandée par Sartorius. Un volume de 400 µL de solution protéique est soumis à plusieurs centrifugations de manière à obtenir un volume de 50 µL dans le système Vivaspin 500. Le système Vivaspin 15 a été utilisé avec un volume d'échantillon de 3 mL permettant d'obtenir 250 µL d'échantillon concentré après centrifugation.

Dialyse

Les échantillons ont été dialysés à travers des membranes SpectrumLabs (Spectra/Por 1 - 6-8 kDa). Les extraits protéiques issus d'*E. coli* sont dialysés contre le tampon 4 (Tableau 26) tandis que ceux issus de *S. cerevisiae* contre le tampon 5 (Tableau 25). Des volumes d'extraits de 3 à 5 mL ont été introduits dans des boudins de dialyse puis mis sous agitation à 4°C durant 24 heures dans un bécher contenant 1 L de la solution de dialyse.

6.2. Recherche d'activité

6.2.1. Test d'hydrolyse de substrat gélosé

L'activité lipolytique des extraits protéiques des levures ou des bactéries modifiées a été testée sur des émulsions gélosées de tributyrine (10 %), de trioléine (2 %), du Tween 20 (2 %) et du Tween 80 (2 %). Les substrats sont préparés dans du tampon phosphate de sodium 20 mM ou 50 mM, supplémenté de 1,5 % (p/v) d'agar. Les solutions sont émulsionnées par sonication puis chauffées jusqu'à ébullition avant d'être coulées en boîtes de Petri ou en microplaques. Après séchage, l'extrait protéique est déposé et incubé à 30 °C et à 37 °C. La présence d'une activité lipolytique extracellulaire doit se traduire par l'apparition d'une zone claire résultant de l'hydrolyse des substrats émulsionnés (pour la tributyrine). La visualisation de la région hydrolysée peut être accentuée en présence de bleu victoria (pour le Tween 20 et 80) ou de rhodamine (pour la trioléine). Pour chaque milieu et condition testés, un témoin positif est réalisé avec une enzyme lipolytique commerciale, la lipase F (Amano Enzyme Europe) extraite de *Rhizopus oryzae*. Des témoins négatifs sont réalisés en

utilisant la construction *pYES2.1-V5-His-LacZ* pour la levure et en utilisant une culture non induite pour la bactérie.

La recherche d'activité sur les extraits protéiques d'*IgTeCe* a également été réalisée sur milieu solide contenant des émulsions de tributyrine (2 % p/v) par zymogramme. Le protocole d'électrophorèse en gel de polyacrylamide est appliqué, sans l'étape de dénaturation des extraits protéiques par la température. Après migration, le SDS est éliminé par lavage du gel deux fois 15 min avec un tampon Tris HCl 20 mM pH 6,4 contenant 0,08 % (p/v) de NaOH, 0,25 % (p/v) d'acide maléique et 25 % (v/v) d'isopropanol et une fois 15 min avec le même tampon mais dépourvu d'isopropanol. Le gel est ensuite déposé sur une boîte de Petri contenant le milieu solide décrit plus haut. L'incubation est réalisée à 37°C jusqu'à apparition de zones de lyse.

6.2.2. Test d'hydrolyse de substrat chromophore

La détection de l'activité lipolytique a été également réalisée par des tests colorimétriques. Différents esters de para-nitrophényle synthétiques ont été testés: le para-nitrophényle butyrate C4 (p-NPB), octanoate C8 (p-NPO), laurate C12 (p-NPL), palmitate C16 (p-NPP). L'activité d'une enzyme lipolytique provoque un clivage de la liaison ester et libère l'acide correspondant et le para-nitrophénol de couleur jaune mesurable par spectrophotométrie à 410 nm.

Le test d'activité utilise les différents esters de para-nitrophényle à 16,5 mM préparés dans l'isopropanol. Le mélange réactionnel est composé d'1 volume de cette solution de substrat et de 9 volumes de tampon Tris-HCl 50 mM (pH 8,0) contenant 0,4% (v/v) de Triton X-100 et 0,1% (p/p) de gomme arabique (Pencreac'h and Baratti, 1997). Le mélange réactionnel est placé à 37 °C. La réaction démarre avec l'ajout de l'enzyme (Lipase F en tant que contrôle positif ou échantillon protéique) préparé dans du tampon Tris HCl 50 mM (pH 8,0).

6.3.Composition des lipides

6.3.1. Extraction des lipides totaux

L'extraction est réalisée sur le culot d'une culture de 50 mL d'*E. coli*. Celui-ci est resuspendu dans un mélange contenant 200 µL d'eau osmosée, 2 mL de méthanol et 1 mL de chloroforme. Pour la quantification, au lieu d'utiliser 1 mL de chloroforme, 1 mL de solution chloroformique d'étalon interne est ajouté. Après broyage dans un tube de dounce, la suspension est incubée toute la nuit sous agitation pour extraire les lipides. La phase inférieure chloroformique est récupérée dans un tube à méthyler. Le chloroforme est ensuite évaporé à sec sous flux d'azote.

6.3.2. Saponification et méthylation

Les acides gras doivent être sous forme d'esters méthyliques afin de les rendre plus volatils et d'être plus facilement détectables par chromatographie en phase gazeuse (CPG). La première étape est une saponification c'est-à-dire une hydrolyse des esters en acides gras libres et alcool. Elle est réalisée en ajoutant à l'échantillon 1 mL de soude méthanolique 0,5 M. Le mélange est chauffé pendant 15 min à 80 °C dans un bain à sec. La seconde étape est une méthylation, réalisée après refroidissement, par ajout de 2 mL de solution méthanolique de trifluorure de Bore (BF₃ à 14 %, Sigma-aldrich™). Après incubation à 80°C pendant 20 min et refroidissement, 2 mL d'isooctane (triméthyl-2,2,4-pentane) sont ajoutés pour réaliser l'extraction des acides gras méthylés. Après la séparation des phases, la phase supérieure d'isooctane est transférée dans un autre tube à méthyler. Cette extraction est répétée deux fois. La phase supérieure d'isooctane est rincée 3 fois avec 4 mL d'eau distillée afin d'éliminer les traces de BF₃ et de soude. Il est vérifié, à l'aide d'un papier pH, que le pH de la dernière eau de rinçage est proche de la neutralité. L'extrait est séché par passage sur une mini colonne (pipette Pasteur) contenant un morceau de laine de verre et du sulfate de sodium anhydre. L'échantillon est alors prêt, il peut être stocké à -20°C avant d'être analysé par CPG.

6.3.3. Analyse par chromatographie en phase gazeuse des esters méthyliques

L'appareil utilisé est le Focus GC ThermoScientific™ muni d'un injecteur split/splitless utilisé en mode split (seule une partie du volume injecté est analysée) à 250 °C, d'un détecteur à ionisation de flamme et d'une colonne capillaire CP SiL 88 de 25 m x 0,25 mm x 0,2 µm. Les injections sont réalisées à l'aide d'un injecteur automatique AI 3000. L'azote est utilisé comme gaz vecteur avec un débit de 1 mL/min. La température du détecteur à ionisation de flamme est fixée à 280 °C. Après l'injection, la température du four est maintenue pendant 4 min à 120 °C puis un gradient de température est appliqué selon une rampe de 120 °C à 220 °C, à raison de 6°C / min. Le ratio du split est fixé à 15 soit un débit de fuite de 15 mL/min. Le volume d'échantillon injecté est de 3 µL et la température de l'injecteur est fixée à 250 °C. L'identification des acides gras est effectuée par comparaison avec le temps de rétention d'esters méthyliques d'acides gras purs analysés dans les mêmes conditions. La quantification de chaque composé est définie par l'aire de son pic (intégrée et normalisée). Afin de quantifier les acides gras, l'acide arachidique C20:0 a été utilisé comme étalon interne à 500 µg/mL. Au lieu d'utiliser 1 mL de chloroforme, 1 mL de solution chloroformique d'étalon interne est ajouté. La concentration d'étalon interne utilisée est estimée à partir d'une gamme d'étalonnage du C20:0.

7. Méthodes bioinformatiques

Des outils de biologie assistée par ordinateur ont été utilisés, lors de cette étude, en local au laboratoire ou via une plateforme en ligne. Les bases de données de références ont été décrites dans le chapitre 1 au paragraphe 2.3.1.

7.1. Manipulations et outils de comparaison de séquences

Les manipulations des séquences génomiques issues de travaux de séquençage ou de bases de données ont été effectuées d'une part à l'aide du logiciel ApE[®]. Ce logiciel a permis de visualiser des séquences nucléotidiques de plasmides et de fragments séquencés (<https://www.projet-plume.org/fiche/ape>). Le logiciel Serial Cloner[®] a d'autre part été employé pour la construction de vecteurs et l'alignement de résultats de séquençage par rapport à des séquences attendues (http://serialbasics.free.fr/Serial_Cloner.html). Ce logiciel a également été utilisé pour la validation de cadres ouverts de lecture et la vérification d'intégrité de gène après séquençage. Enfin, SerialCloner[®] a permis de réaliser la traduction in silico de manière à travailler à partir de séquences protéiques. Le logiciel AmplifX[®] a été appliqué pour la prédiction et la simulation de réactions PCR (<http://crn2m.univ-mrs.fr/recherche/equipe-t-brue/jullien-nicolas/programmation/amplifx/>).

De nombreux outils sont disponibles en ligne sur des serveurs de plateformes bioinformatiques ou directement sur les sites des laboratoires développant ces outils spécifiques. L'alignement de séquences (protéiques et génomique) du gène candidat *IgTeCe* a été réalisé contre des séquences connues grâce à l'outil BLAST. Ses variantes PHI-BLAST et DELTA-BLAST ont été utilisées sur les séquences protéiques déduites grâce à SerialCloner[®]. Ces algorithmes ont permis d'affiner des identifications en fonction de l'intégrité de la séquence mais également de la position du site actif à l'intérieur de celle-ci (<http://BLAST.ncbi.nlm.nih.gov/BLAST.cgi>). Ensuite, les alignements de séquences protéiques ont été analysés par l'installation en local du logiciel Clustal-X2[®] (<http://www.clustal.org/>). La suite logicielle PipeAlign a été également utilisée pour automatiser la comparaison des séquences protéiques (Plewniak *et al.*, 2003). De plus, divers outils de la suite de logiciels EMBOSS ont été utilisés pour déterminer les signaux présents au sein de la séquence (bips.u-strasbg.fr/EMBOSS/).

7.2. Profils physico-chimiques des protéines et prédictions de structures

La prédiction des propriétés physico-chimiques des protéines a été réalisée grâce à plusieurs logiciels. La présence de signaux peptidiques dans la séquence protéique a été évaluée par le logiciel SignalP (http://bioinformatics.leeds.ac.uk/prot_analysis/Signal.html). La prédiction de formation de ponts disulfure a été réalisée par l'algorithme DISULFIND (Ceroni *et al.*, 2006). La masse moléculaire et le point isoélectrique ont été déterminés sur le site Expasy (www.expasy.org) (Artimo *et al.*, 2012). Les sites de glycosylation sont prédits à l'aide de NetNGlyc, NetOGlyc et NetCGlyc (<http://www.cbs.dtu.dk/services/>). Un second algorithme est utilisé pour la caractérisation des sites d'O-glycosylation à l'aide de GlycoEP (<http://www.imtech.res.in/raghava/glycoep/index.html>).

Les structures secondaires prédictives ont été établies par les logiciels YASPIN et Jpred (Lin *et al.*, 2005; Cole *et al.*, 2008). Les structures prédictives ont été établies à l'aide de Modeller (Eswar *et al.*, 2001). La visualisation et l'export sous forme d'image de structures tridimensionnelles ont été réalisés par le logiciel PyMol (<http://www.pymol.org/>).

7.3. Analyses et optimisation de prises de vue expérimentales

Deux logiciels d'analyse de photos ont été utilisés pour la prise de vue et l'optimisation des gels et membranes de transferts : le logiciel ImageJ™ (<http://fiji.sc/>) et le logiciel Quantity One® (<http://www.bio-rad.com/prd/en/US/LSR/SKU/170-9600/Quantity-One-1-D-Analysis-Software>).

CHAPITRE III :

RESULTATS ET DISCUSSION

CHAPITRE III : RESULTATS ET DISCUSSION

1. Préambule

Ce chapitre présente les résultats obtenus au cours de cette thèse. Une étude préliminaire avait permis l'identification et la caractérisation de dix séquences pouvant coder des enzymes lipolytiques. Ces précédents travaux, décrits au paragraphe 2.4.2 du chapitre I, avaient permis d'aboutir au clonage de séquences codantes pour une expression dans la levure *Saccharomyces cerevisiae*.

Les travaux effectués durant ma thèse s'inscrivent dans la continuité de ce projet en s'intéressant à l'expression hétérologue de deux séquences codantes, *IgEst3* et *IgTeCe*. Elles possèdent toutes les deux le pentapeptide consensus GXSXG. L'objectif est la caractérisation des fonctions enzymatiques des protéines exprimées à partir de ces deux séquences codantes.

Ce chapitre est divisé en cinq parties. La première partie des résultats est consacrée à l'analyse bioinformatique des deux séquences protéiques déduites. La seconde partie présente les résultats obtenus lors de l'expression hétérologue d'*IgEst3* et d'*IgTeCe* dans *Saccharomyces cerevisiae*. La troisième partie est consacrée aux résultats obtenus lors de l'expression d'*IgTeCe* dans l'hôte bactérien *Escherichia coli*. Une quatrième partie est consacrée à la caractérisation de l'activité d'*IgTeCe*. Enfin, la cinquième partie est une discussion générale de l'ensemble de ces travaux de thèse.

2. Etudes *in silico* des deux séquences codantes d'*Isochrysis galbana*

2.1. Etude d'*IgEst3*

Les résultats et prédictions obtenus par les outils bio-informatiques (caractéristiques physico-chimiques, modifications post-traductionnelles) ont une importance dans le choix des conditions d'extraction, d'isolement et de détection de la protéine. Au niveau protéique, l'étude *in silico* va également permettre de formuler des hypothèses de travail pour la fonction enzymatique et le choix des tests d'activité (recherche de séquences homologues, modélisation de structures).

La séquence codante *IgEst3* est une séquence de 990 pb codant une protéine putative de 329 acides aminés. La comparaison de la séquence génomique avec l'ADNc indique que le gène *IgEst3* semble contenir un intron de 90 paires de base entre les positions 369 et 459. La figure 25 présente la séquence codante et la séquence protéique déduite d'*IgEst3*. La séquence incomplète disponible dans la base de données TBestDB (cluster ISL00006895), est surlignée en bleu. Le pentapeptide GFSQG,

correspondant à la séquence consensus GX SXG, est indiqué en vert. La sérine centrale du pentapeptide est susceptible d'appartenir à la triade catalytique.

```

-----|-----|-----|-----|-----|-----|
1  atgagcctggtcatgctcgacacttcccctcaggccagggtgacccagcaaggacg 60
1  M S P G H A R D T S P Q A R V H P A R T 20

-----|-----|-----|-----|-----|-----|
61  acaggaggaccaatggggcggcagctacgaaaacgacgcctgcggcagggttc 120
21  T G G P M G R Q L R K R R L R A G F W Q 40

-----|-----|-----|-----|-----|-----|
121  gggatgcgtgagggcgaaaatgagaagggacctctcgagtttgagccgatggagagccg 180
41  G M R E G E N E K G P L E F E P M E S P 60

-----|-----|-----|-----|-----|-----|
181  tggaccgctgcagtgatttggctccacggcttggcgatgaacccgaggcatgggcccga 240
61  W T A A V I W L H G F G D E P E A W A E 80

-----|-----|-----|-----|-----|-----|
241  ttcattggcggcagataggcgagcaaggccgctcttggcctgggtcttccctcgggccaag 300
81  F M A A D R R A R P S W A W V F P R A K 100

-----|-----|-----|-----|-----|-----|
301  gtactgcgcagccttgctacgggtggcaggagctgaccgcattggcgagttttctggac 360
101  V L P Q P C Y G G Q E L T A W A S F L D 120

-----|-----|-----|-----|-----|-----|
361  gctgaggtcatccgccgggtggcgctgattacgactcgagcgcaacaacttcgccaag 420
121  A E V I R P G G A D Y D S S A N N F A K 140

-----|-----|-----|-----|-----|-----|
421  acagtcgctctcatccatcgacgcttgatgcggtacgtgcacg 480
141  T V A L I H R S V D A V R A R H N L P A 160

-----|-----|-----|-----|-----|-----|
481  gacgcgctcgccattgcccgttctcgcagggcgctgccgcgcgcgcgcgcgcctc 540
161  E R V A I A G F S Q G A A A A A A A A L 180

-----|-----|-----|-----|-----|-----|
541  aattaccctgagcgtctcgccgcatcggttgcctccggctggctgctcaaaagcgtct 600
181  N Y P E R L G G I G L L S G W L L K A S 200

-----|-----|-----|-----|-----|-----|
601  cgccagacgtgagccggcgcgagtcgagctgtggtccaggcgcatgcttcttcgtg 660
201  R Q T V S R A R S P S G P G A F F V 220

-----|-----|-----|-----|-----|-----|
661  agtcacggaacggccgacacccaggctcgcttgaatgcgcagaggaatccacgcgactt 720
221  S H G T A D T Q V A F E C A E E S T R L 240

-----|-----|-----|-----|-----|-----|
721  ctgcgcgagagtggggcatcggtggtcgccgcgccttccaaaacttgagcacgaagct 780
241  L R E S G A S V V A R A F P K L E H E A 260

-----|-----|-----|-----|-----|-----|
781  ggggccaatgcagctgtacgtgacgtcatgtgttcttgatgactgcttgtgtgtccg 840
261  G A N A A V P D V M F L D D L A P 280

-----|-----|-----|-----|-----|-----|
841  gtggcagtcctgctcgtgaggaacggcgccaaccccgcaagcccagggtgtaccat 900
281  V A V R A R E E T G A N P A Q A Q V Y H 300

-----|-----|-----|-----|-----|-----|
901  accagcgcgaggagaagggtcggcacgccctgagcttggctgcttacggagaagcaaaaag 960
301  T S A G E G S A R P E L G L R R S K K 320

-----|-----|-----|
961  gcgcgtgacgtgaccgtgacgtgcattag 990
321  A R D V T V I V H * 329

```

Figure 25: Séquences codante et protéique d'*IgEst3*.

Le pentapeptide correspondant à la séquence consensus GX SXG est surligné en vert et composé comme suit : GFSQG.

La séquence correspondant à la séquence ISL00006895 issue de la base de donnée EST apparait en bleu.

2.1.1. Prédiction des caractéristiques d'*IgEst3*

Recherche d'un signal peptidique

Cette recherche est indispensable pour identifier dans la séquence protéique d'éventuels signaux peptidiques consensus reconnus dans les organismes procaryotes et eucaryotes. L'outil utilisé, SignalP, présente les résultats sous forme de trois valeurs : C, S et Y. La valeur C correspond à la prédiction des signaux de clivage, la valeur S correspond à la prédiction de signaux peptidiques caractérisés et la valeur Y correspond à la prédiction de signaux de clivage en fonction de la valeur S.

Le graphique obtenu à l'aide de cet algorithme est présenté en annexe 1. Aucune des trois valeurs n'est significative pour la protéine *IgEst3* et aucun signal peptidique n'est détecté. La protéine *IgEst3* ne présente pas de caractéristiques pour être orientée vers un compartiment cellulaire donné.

Evaluation de la masse moléculaire et du point isoélectrique

La séquence d'*IgEst3* code une protéine de masse moléculaire théorique de 35,41 kDa et de pI théorique de 8,66.

Prédiction de la formation de ponts disulfures

Le logiciel de prédiction a identifié 6 acides aminés cystéines (présentés en rouge sur fond gris sur la figure 25) susceptibles de former par paire des ponts disulfures. Cependant, le logiciel Disulfind indique une faible probabilité de formation de ces structures.

Prédiction de la présence de sites de glycosylation

Cette prédiction a pour objectif d'évaluer si la protéine *IgEst3* sera ou non glycosylée lors de l'expression dans la levure. La recherche de ces sites de glycosylation a été entreprise par trois logiciels, NetNGlyc, NetOGlyc et GlycoEP, dédiés à des protéines eucaryotes.

La prédiction est basée sur deux critères: la présence de signaux peptidiques capables d'orienter la protéine dans les compartiments de glycosylation et la présence de sites reconnus par les systèmes de glycosylation.

Les sites consensus de N-glycosylation sont composés d'une suite d'acides aminés : Asn-X-Ser/Thr où l'acide aminé X ne peut pas être une proline. Pour chaque nouvelle analyse, l'algorithme « NetNGlyc » recherche dans un premier temps les signaux peptidiques permettant l'orientation vers les compartiments de glycosylation. Dans un second temps, il recherche le 1^{er} acide aminé du site (Asn) et détermine ensuite si les deux acides aminés suivants l'asparagine sont compatibles avec le site consensus. Les résultats obtenus, présentés en annexe 2, indiquent l'absence de signaux peptidiques

et l'absence de sites de N-glycosylation. Ces deux résultats rendent improbables la formation de N-glycosylation pour *IgEst3*.

La prédiction de la formation d'O-glycosylation a également été étudiée. Elle est obtenue par association des résultats de deux algorithmes « NetOGlyc » et « GlycoEP ». Le premier algorithme a une matrice de comparaison fixe, alors que le second calcule une matrice de comparaison par alignement de séquences.

La présence de sites d'O-glycosylation est significative et de nombreux résidus compatibles ont été détectés. Les positions potentiellement glycosylées détectées par chaque algorithme sont présentées en annexe 3. Toutefois, les signaux peptidiques permettant l'orientation vers les compartiments cellulaires de glycosylation sont absents. Seule une orientation de la protéine vers ces compartiments permet sa glycosylation. En conclusion, la formation de glycosylations dans la levure apparaît peu probable.

2.1.2. Prédications de la fonction enzymatique par recherche d'homologies

La séquence protéique d'*IgEst3* a été comparée aux séquences répertoriées dans les banques de données publiques en utilisant d'une part l'algorithme BLASTP et d'autre part le logiciel automatisé PipeAlign. Les résultats de score, de pourcentage d'identité et de pourcentage de similarité permettent d'évaluer l'homologie entre deux séquences. Le pourcentage d'identité correspond aux résidus strictement conservés tandis que le pourcentage de similarité correspond aux résidus de mêmes propriétés biochimiques. Cette démarche permet d'émettre une hypothèse sur la fonction enzymatique pour deux protéines présentant des homologies de séquence.

Recherche par l'algorithme BLASTP

Le tableau 35 présente les différents programmes d'alignement retenus pour rechercher des séquences homologues à *IgEst3*. L'objectif est de pouvoir combiner les différents résultats d'alignement et de prédire une fonction enzymatique.

Méthodologie	Principe	Particularités
BLASTP	Comparaison simple de la séquence candidate aux séquences protéiques d'une base de données. La hiérarchie des résultats est fonction d'une matrice standard.	Identification des séquences proches sans forcément identifier une fonction.
DELTA-BLAST	Construction d'une matrice spécifique à la séquence candidate. Comparaison des domaines conservés entre les séquences.	Evolution importante de l'alignement de séquences protéiques.
PHI-BLAST	Conservation du principe du BLASTP, cependant les séquences comparées doivent contenir un motif protéique bien précis, ici le pentapeptide GX SXG.	Permet de restreindre à certaines familles d'enzymes sans sélectionner une banque particulière.

Tableau 35: Programmes d'alignement

Chaque méthodologie a été utilisée avec une sélection de banques de données afin d'élargir ou de restreindre l'obtention de séquences homologues. Le tableau 36 présente ces différentes banques de données.

Banque de données	Description de la banque de donnée
NR	Regroupe les séquences codantes de la GenBank traduites, les séquences de la PDB, de la SwissProt, de la PIR et de la PRF.
RefSeq	Regroupe strictement les séquences protéiques du projet de séquences de référence de la NCBI.
SwissProt	Regroupe strictement les séquences de référence de la SwissProt issues de la dernière mise à jour.
PDB	Regroupe exclusivement les séquences protéiques associées à des structures biologiquement démontrées de la Protéine Data Bank.

Tableau 36: Banques de données utilisées pour l'alignement de la séquence *IgEst3*.

Les associations Méthodologie/Banque de données ayant permis d'obtenir des séquences homologues informatives pour la prédiction de la fonction d'*IgEst3* sont présentées dans le tableau 37.

Numéro d'alignement	Méthodologie choisie	Banque de données choisie	Organismes choisis	Protéines putatives incluses
1	BLASTP	NR	Tous	Oui
2	BLASTP	RefSeq	Tous	Non
3	BLASTP	RefSeq	Plantes /Algues	Non
4	BLASTP	SwissProt	Tous	Non
5	BLASTP	PDB	Tous	Non
6	DELTA-BLAST	NR	Tous	Non
7	DELTA-BLAST	NR	Plantes	Non
8	DELTA-BLAST	NR	Algues	Non
9	PHI-BLAST	NR	Tous	Non

Tableau 37: Liste des alignements retenus pour la protéine *IgEst3*

Les séquences obtenues ont été annotées selon le principe décrit en annexe 4. Les séquences homologues de chacun des alignements sont référencées en annexe 5.

Pour chaque alignement une valeur de score, un pourcentage d'identité et un pourcentage de similarité sont calculés. Le score est une valeur permettant de traduire la qualité de l'alignement, 0 correspond à une faible qualité de l'alignement tandis que 200 correspond à une qualité très élevée.

Dans la suite de ce paragraphe, chaque alignement de séquences est brièvement décrit. Les séquences homologues sont classées en fonction du score de manière décroissante puis en fonction du pourcentage d'identité et de similarité.

- Alignement 1 : BLASTP/NR.

Cet alignement est réalisé à l'aide de BLASTP sur l'ensemble des banques NR. Il s'agit de l'alignement le plus classique. Le résultat présenté dans le tableau 38 montre que les séquences homologues à *IgEst3* issues de micro-organismes marins, présentent des scores supérieurs à 100 et des

pourcentages de similarité proche de 50 %. Les pourcentages de similarité sont supérieurs aux pourcentages d'identité, proches de 30 %. Deux séquences correspondent à des protéines hypothétiques issues des microalgues *Emiliana huxleyi* [*EhHp1*] et *Guillardia theta* [*GtHp1*]. La troisième séquence [*IdCe1*] est prédite comme une carboxylestérase et issue de la bactérie marine *Idiomarina* sp.. Les trois séquences suivantes présentent un score inférieur à 100 avec des pourcentages d'identité et de similarité proches des séquences précédentes. Seule la séquence homologue [*RfCe1*] issue de la bactérie *Rhodospirillum rubrum* est annotée comme une carboxylestérase. Cette bactérie est retrouvée dans les sédiments de zones de balancement des marées. Les deux autres séquences homologues, l'une issues du champignon *Trichoderma reesei* [*TrHp1*] et l'autre de la larve *Oikopleura dioica* [*OdUpp1*], sont des protéines hypothétiques ou inconnues.

	[<i>EhHp1</i>]	[<i>GtHp1</i>]	[<i>IdCe1</i>]	[<i>TrHp1</i>]	[<i>RfCe1</i>]	[<i>OdUpp1</i>]
Score	138	104	102	97,1	92	94,7
Identité %	37 %	32 %	31 %	35 %	35 %	29 %
Similarité %	51 %	47 %	50 %	46 %	46 %	43 %

Tableau 38: BLASTP d'*IgEst3* contre les séquences des banques NR.

Les séquences retenues lors de cet alignement BLASTP sont majoritairement des protéines hypothétiques ou inconnues (*Hp* ou *Upp*). Cet alignement montre deux séquences homologues correspondant à la fonction carboxylestérase. Pour confirmer ce résultat, la recherche doit être approfondie en changeant d'algorithme et/ou de bases de données. Seules les séquences avec une fonction enzymatique bien annotée seront prises en compte dans la suite de cette étude.

- Alignement 2 : BLASTP/RefSeq

L'alignement de la séquence *IgEst3* avec les séquences de la base de données RefSeq (Tableau 39), a montré que les meilleurs scores sont obtenus avec les carboxylestérases [*IdCe1*] et [*RfCe1*] précédemment alignées par BLASTP/NR. Des scores élevés sont également obtenus avec des séquences putatives de carboxylestérases issues de la bactérie *Taylorella asinigenitalis* [*TaCe1*] et des bactéries du genre *Herbaspirillum* sp. [*HeCe1*].

	[<i>IdCe1</i>]	[<i>RfCe1</i>]	[<i>TaCe1</i>]	[<i>HeCe1</i>]
Score	102	92	91,7	90,5
Identité %	31 %	35 %	33 %	35 %
Similarité %	50 %	46 %	48 %	47 %

Tableau 39: BLASTP d'*IgEst3* contre les séquences de la banque RefSeq.

L'alignement BLASTP/RefSeq met en évidence des homologies de séquences *IgEst3* avec des séquences de carboxylestérases procaryotes issues notamment de deux bactéries marines.

- Alignement 3 : BLASTP/RefSeq/Plantes-Algues

Pour poursuivre l'accumulation d'informations et affiner la recherche, l'alignement suivant est réalisé en se focalisant sur des séquences d'algues et de plantes (Tableau 40). Chez les plantes, les meilleurs scores sont obtenus pour des séquences putatives de carboxylestérase d'*Arabidopsis thaliana* [AtCe2], d'acylprotéine thioestérases de la vigne commune *Vitis vinifera* [VvTe1] et de la fraise sauvage *Fragaria vesca* [FvTe1] et [FvTe2]. Une séquence [ZmTe1] identifiée comme thioestérase et issue du maïs est également retenue comme homologue à *IgEst3*. L'analyse des séquences propres aux plantes a donc permis de caractériser la fonction thioestérase comme une nouvelle fonction possible pour *IgEst3*.

L'alignement de séquences issues d'algues donne peu de résultats exploitables. Seule la séquence [NgLp2] affiche un score supérieur à 80. Elle est annotée comme une lysophospholipase issue de la microalgue *Nannochloropsis gaditana*. Cet organisme est également étudié, comme la microalgue *Isochrysis galbana*, pour sa composition lipidique.

	Plantes					Algues
	[AtCe2]	[VvTe1]	[FvTe2]	[FvTe1]	[ZmTe1]	[NgLp2]
Score	83,6	82,8	81,6	81,6	80,5	87,8
Identité %	32 %	33 %	32 %	32 %	33 %	29 %
Similarité %	45 %	42 %	44 %	44 %	45 %	44 %

Tableau 40: BLASTP d'*IgEst3* contre les séquences de la banque RefSeq et issues de plantes et d'algues.

L'alignement 3 nous oriente vers trois fonctions possibles : carboxylestérase, thioestérase et lysophospholipase.

- Alignement 4 : BLASTP/SwissProt

Afin de compléter cette analyse, la séquence *IgEst3* a été comparée à la banque de données SwissProt annotée manuellement. Cette base recoupe à la fois des informations d'activité biologique et des résultats d'expérimentation ou de séquençage de génomes. Les résultats montrent des similarités avec des séquences putatives de thioestérases provenant de levures et de champignons (Tableau 41). En effet, le meilleur score est obtenu avec la séquence [DbTe1] issue de *Debaryomyces hansenii*, levure qui a la particularité d'être tolérante aux faibles températures et résistante aux taux de salinité importants.

	[DbTe1]	[CaTe1]	[UmTe1]	[OcLp1]	[RnLp1]
Score	90,9	81,6	78,6	75,9	74,7
Identité %	27 %	24 %	27 %	28 %	27 %
Similarité %	48 %	44 %	42 %	42 %	42 %

Tableau 41: BLASTP d'*IgEst3* contre les séquences de la banque SwissProt.

Les autres séquences proviennent de la levure *Candida albicans* [*CaTe1*] et du champignon *Ustilago maydis* [*UmTe1*]. La fonction thioestérase semble prépondérante. Cet alignement montre également des similarités avec une lysophospholipase d'*Oryctolagus cuniculus* [*OcLp1*] et une phospholipase de *Rattus norvegicus* [*RnLp1*], avec des scores d'environ 75.

- Alignement 5 : BLASTP/PDB

En comparant avec des séquences caractérisées provenant de la banque de données PDB, peu de séquences s'alignent avec *IgEst3*. La distance phylogénétique est plus importante, ce qui peut expliquer des valeurs de score inférieures à 80. Les pourcentages d'identité sont proches de ceux obtenus lors de l'alignement précédent (Tableau 42). La première séquence [*FtCe1*] est une séquence de carboxylestérase issue de la bactérie *Francisella tularensis*. La seconde séquence [*HsTe1*] est une séquence de thioestérase humaine (hAPT1). La troisième séquence [*PfCe2S*], est une séquence de carboxylestérase issue de *Pseudomonas fluorescens*. Cette structure sert de référence pour l'étude des conformations de carboxylestérases. La quatrième séquence, homologue à la séquence d'*IgEst3* est la séquence de la lysophospholipase 1 humaine [*HsLYPL1S*].

	[<i>FtCe1</i>]	[<i>HsTe1</i>]	[<i>PfCe2S</i>]	[<i>HsLYPL1S</i>]
Score	70,9	67,4	57,8	55,8
Identité %	22 %	28 %	26 %	26 %
Similarité %	42 %	43 %	41 %	41 %

Tableau 42: BLASTP d'*IgEst3* contre les séquences de la banque PDB.

L'alignement 5 a permis l'analyse de certains domaines conservés. En effet, les positions de la séquence consensus GX SXG et des acides aminés composant la triade catalytique sont définies dans les structures, et ont ainsi pu être prédites dans *IgEst3*. Le positionnement hypothétique de ces sites est indiqué en figure 26. La sérine présente dans le pentapeptide GX SXG est en position 169 (S¹⁶⁹), la triade catalytique se complète par les acides aminés D²²⁶ et H²⁵⁸.

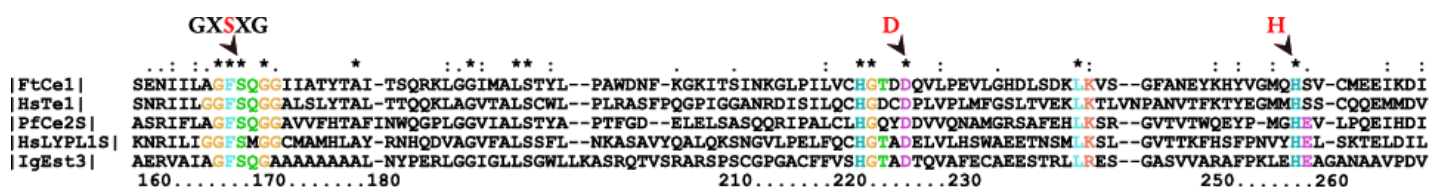


Figure 26: Alignement des séquences homologues obtenues dans la banque PDB.

Les acides aminés signalés en rouge correspondent théoriquement aux résidus de la triade catalytique S, D, H.

- Alignement 6 : DELTA-BLAST/NR

L'algorithme BLASTP a permis d'identifier les fonctions putatives d'*IgEst3* soit comme carboxylestérase, soit comme thioestérase. Pour affiner ces prédictions de fonction et les hiérarchiser,

des alignements complémentaires ont été réalisés avec DELTA-BLAST. Cette recherche est basée sur l'identification de domaines conservés afin de pré-sélectionner les séquences à aligner. Les résultats obtenus par DELTA-BLAST sont présentés dans le tableau 43. Les trois premières séquences sont issues d'organismes du milieu marin. Deux sont d'origine procaryote. Elles correspondent à une séquence putative de thioestérase issue de la bactérie *Glaciecola chathamensis* [*GcTe1*] et à une séquence putative de carboxylestérase issue de l'annotation du génome de la bactérie marine *Kangiella koreensis* [*KkCe1*]. La séquence [*UbCe1*] s'apparente à une séquence de carboxylestérase bactérienne et provient d'une étude de métagénomique réalisée sur des sédiments marins. Deux séquences homologues viennent compléter ce résultat, [*ApTe1_2*], une séquence identifiée comme une thioestérase issue du puceron vert *Acyrtosiphon pisum* et [*PsCe1*], une séquence putative de carboxylestérase issue de *Pseudomonas synxantha*.

	[<i>GcTe1</i>]	[<i>KkCe1</i>]	[<i>UbCe1</i>]	[<i>ApTe1_2</i>]	[<i>PsCe1</i>]
Score	116	111	111	110	109
Identité %	29 %	23 %	24 %	23 %	24 %
Similarité %	42 %	39 %	44 %	40 %	37 %

Tableau 43: DELTA-BLAST d'*IgEst3* contre les séquences des banques NR.

- Alignement 7 : DELTA-BLAST/NR/Plantes

L'alignement DELTA-BLAST a également été réalisé spécifiquement pour les plantes. Malgré un score élevé, la séquence [*RcOSBS*] n'a pas été retenue car ce n'est pas une hydrolase à sérine. Elle correspond à l'enzyme de synthèse de la menaquinone (Vitamine K-2) dans le plant de ricin (Tableau 44). Les meilleurs scores sont obtenus avec des séquences putatives de carboxylestérases [*AlCe1*] et/ou phospholipase [*JcPlCe1*] et de thioestérase [*RcTe1*], [*AtSOBER1*]. Cette dernière, issue d'*Arabidopsis thaliana* est une séquence orthologue aux séquences [*HsTe1*] et [*HsLYPLIS*] analysées dans l'alignement 5.

	[<i>RcOSBS</i>]	[<i>AlCe1</i>]	[<i>JcPlCe1</i>]	[<i>RcTe1</i>]	[<i>AtSOBER1</i>]
Score	92	91,7	91,3	90,5	88,2
Identité %	16 %	23 %	22 %	22 %	27 %
Similarité %	30 %	36 %	32 %	33 %	40 %

Tableau 44: DELTA-BLAST d'*IgEst3* contre les séquences des banques NR issues de plantes.

- Alignement 8 : DELTA-BLAST/NR/Algues

L'algorithme DELTA-BLAST a par la suite été réalisé spécifiquement pour les séquences issues d'algues (Tableau 45). Les valeurs des scores et les pourcentages d'identité et de similarité sont comparables à celles du DELTA-BLAST avec des séquences de plantes.

Pour cet alignement, il faut distinguer trois groupes de séquences homologues. Le premier correspond aux microalgues chlorophycées et contient la séquence homologue $|OtLp1|$, une séquence putative de lysophospholipase, issue d'*Ostreococcus tauri* et la séquence $|BpCe1|$, une séquence putative de carboxylestérase issue de *Bathycoccus prasinos*. Le second groupe contient une séquence putative de carboxylestérase de la macroalgue rhodophycée *Chondrus crispus*, la séquence $|CcCe1|$. Le troisième groupe correspond aux microalgues *Chromalveolata* (incluant *Isochrysis galbana*) et contient la séquence $|NgLp3|$, une séquence putative de lysophospholipase, issue de *Nannochloropsis gaditana*.

	$ OtLp1 $	$ BpCe1 $	$ CcCe1 $	$ NgLp3 $
Score	100	95,1	92	87,4
Identité %	26 %	22 %	21 %	26 %
Similarité %	37 %	38 %	32 %	40 %

Tableau 45: DELTA-BLAST d'*IgEst3* contre les séquences des banques NR issues d'algues.

- Alignement 9 : PHI-BLAST/NR

En parallèle de ces analyses, un alignement de séquences a été réalisé en tenant compte de la présence du pentapeptide consensus GX SXG. Cet alignement a été effectué par l'utilisation de l'algorithme PHI-BLAST. La plupart des séquences obtenues par cet alignement ont déjà été identifiées lors des alignements de séquences avec BLASTP (sur fond gris, tableau 46). Les valeurs de score ont permis de s'intéresser à trois séquences supplémentaires. La première séquence $|TrPp1|$ est une séquence sans fonction enzymatique prédite issue d'un champignon filamenteux *Trichoderma reesei*. La seconde séquence $|HeCe1|$ est une séquence putative de carboxylestérase issue des bactéries *Herbaspirillum* sp., liées à la croissance de plantes herbacées. La troisième séquence $|HfCe1|$ est une séquence putative de carboxylestérase issue d'*Herbaspirillum frisingense*.

	$ EbHp1 $	$ GtHp1 $	$ IdCe1 $	$ RfCe1 $	$ TrPp1 $	$ HeCe1 $	$ HfCe1 $
Score	141	83,1	81,5	77,3	77	73,5	73,1
Identité %	35 %	32 %	31 %	35 %	34 %	32 %	32 %
Similarité %	47 %	47 %	51 %	46 %	47 %	45 %	44 %

Tableau 46: PHI-BLAST d'*IgEst3* contre les séquences des banques NR sans restriction d'organismes, l'alignement tient compte de la présence du pentapeptide GX SXG.

- Recherche de séquences homologues dans les banques de données ESTHER, MELDB et LED

La recherche de séquences homologues au sein de ces trois banques a été réalisée par BLASTP. Le résultat pour les banques LED et ESTHER est identique. La séquence présentant le plus d'homologie est la séquence de carboxylestérase $|RfCe1|$ décrite aux alignements 1 et 2. L'alignement de la séquence *IgEst3* avec la base de données MELDB a eu pour meilleur résultat la séquence

[BbCe1] issue de la bactérie *Bordetella bronchiseptica*. Cette séquence correspond à la fonction carboxylestérase. L'alignement avec les séquences de lipases de la base MELDB n'a pas donné de résultats significatifs (score ≤ 30).

Recherche avec le logiciel PipeAlign

Cette stratégie se base sur l'utilisation du logiciel PipeAlign qui est une suite de logiciels d'analyse dont le résultat de chaque étape sert de base pour l'étape suivante. Cette approche permet d'obtenir une classification fonctionnelle. A l'issue de l'ensemble des traitements, deux groupes ont été constitués (Figure 27). Le groupe A, contenant notamment la séquence *IgEst3*, est principalement composé de protéines non annotées. Cependant, la fonction thioestérase est la fonction enzymatique la plus largement représentée. Dans le groupe B, constitué de séquences dont l'identité par rapport à *IgEst3* est plus faible, les fonctions enzymatiques phospholipase et carboxylestérase sont les plus représentées. De nombreuses séquences sont doublement annotées avec la fonction carboxylestérase et la fonction phospholipase. L'analyse de ces résultats oriente l'identité d'*IgEst3* vers une fonction de thioestérase plutôt qu'une fonction de carboxylestérase. En outre, il est intéressant de noter l'importante proportion de séquences proches d'*IgEst3* encore non annotées ou identifiées.

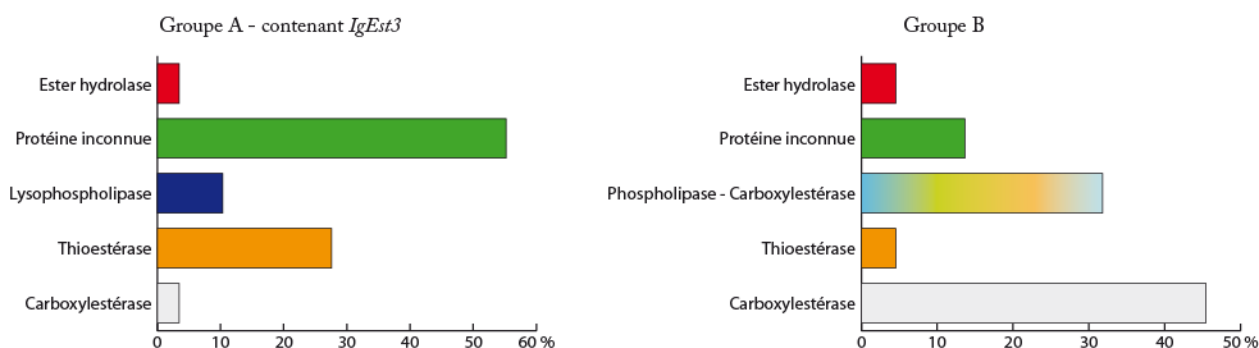


Figure 27: Composition des deux groupes d'enzymes obtenus par le logiciel PipeAlign

2.1.3. Prédiction de l'organisation structurale

La recherche de l'organisation structurale s'est d'abord portée sur l'organisation de la structure secondaire qui peut être déterminée par deux méthodes.

La première méthode est un calcul direct utilisant l'algorithme YASPIN (Figure 28 A). Ce dernier utilise l'alignement de séquences de manière à établir une matrice de calcul. Des séquences dont la structure a été caractérisée sont alors sélectionnées à partir de cette matrice et utilisées pour définir l'emplacement des différents domaines. La schématisation en figure 28 A montre en rouge les

structures de type hélice, en vert les structures de type feuillet et enfin en bleu les liaisons non structurées.

La seconde méthode utilise la prédiction de la structure tridimensionnelle pour en donner un aperçu sous forme de structure secondaire (Figure 28 B). La structure tridimensionnelle a été obtenue par modélisation. Dans la figure 28 B, les cylindres schématisent les hélices, les flèches les feuillets et les lignes les liaisons non structurées.

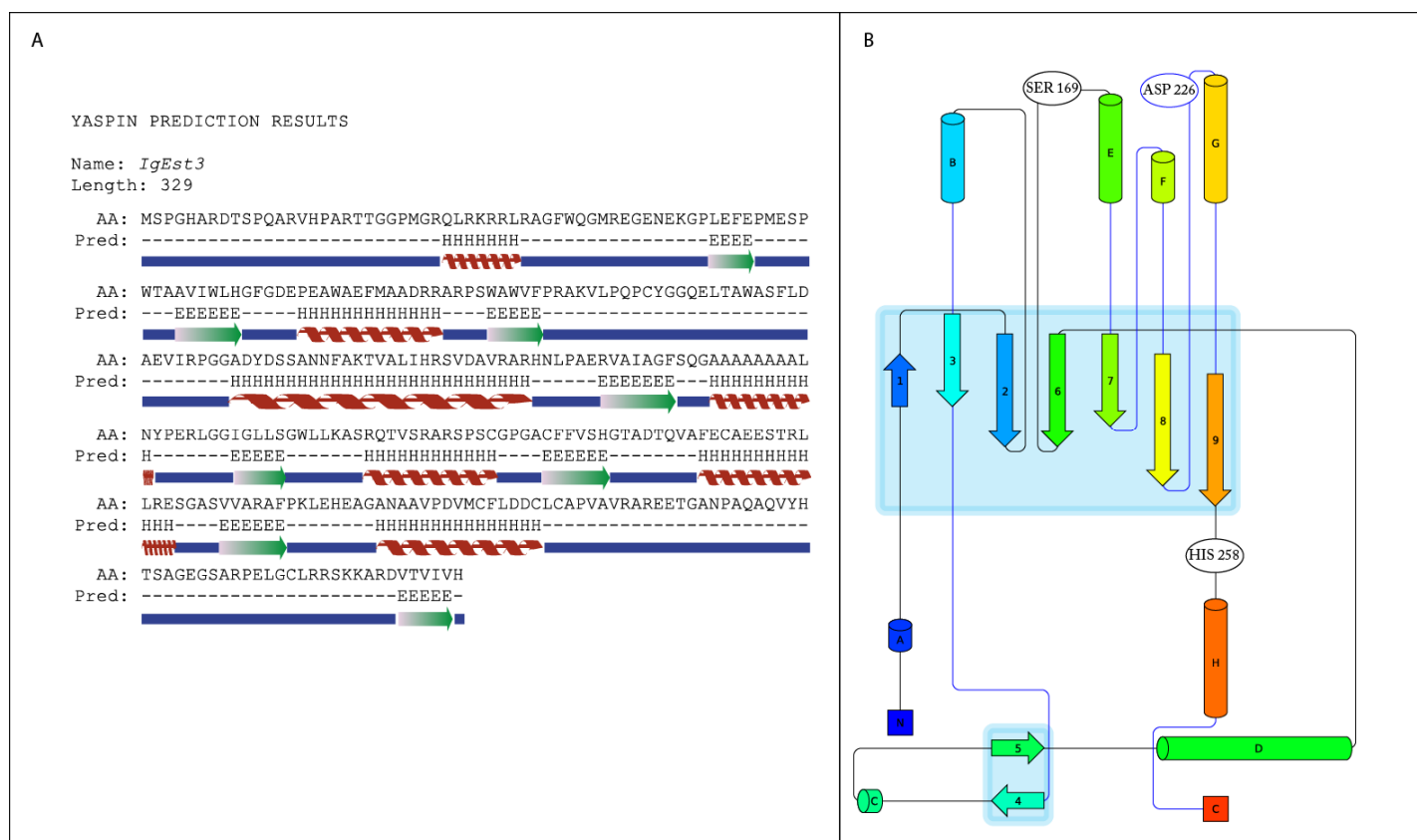


Figure 28: Prédiction des structures secondaires d'*IgEst3*.

A - Représentation graphique des différentes structures composant la séquence (algorithme YASPIN).

B- Structure secondaire déduite de la structure tridimensionnelle modélisée pour la séquence *IgEst3*.

Dans les deux cas, la modélisation des structures secondaires montre une succession d'hélices α et de feuillets β correspondant à l'organisation des hydrolases α/β . Les positions des acides aminés composant la triade catalytique ayant été théoriquement identifiées, il est possible de les localiser sur la modélisation structurale. Chacun des acides aminés est retrouvé à la jonction entre une structure de type feuillet et une structure de type hélice. La sérine est située entre le feuillet 6 et l'hélice E, l'acide aspartique entre le feuillet 8 et l'hélice G et enfin l'histidine, entre le feuillet 9 et l'hélice H (Figure 28 B).

Globalement, les deux modèles prédits convergent. Cependant, l'algorithme YASPIN ne prédit ni l'hélice « C » ni les feuillets « 4 et 5 » mais prédit un feuillet en C-terminal absent de la structure secondaire déduite de la structure tridimensionnelle.

Une modélisation tridimensionnelle a également été réalisée à l'aide du logiciel Modeller à partir des séquences résultant de l'alignement 5 BLASTP/PDB. La figure 29 présente la structure de la triade catalytique. Les flèches plates représentent les feuillets, les structures rubans représentent les hélices. Les acides aminés de la triade catalytique sont représentés avec leurs chaînes latérales (en rouge). Leur positionnement semble cohérent par rapport à la structure de la carboxylestérase de *Francisella tularensis* (à droite). La composition de la séquence d'*IgEst3* semble donc compatible avec les structures tridimensionnelles d'hydrolases connues.

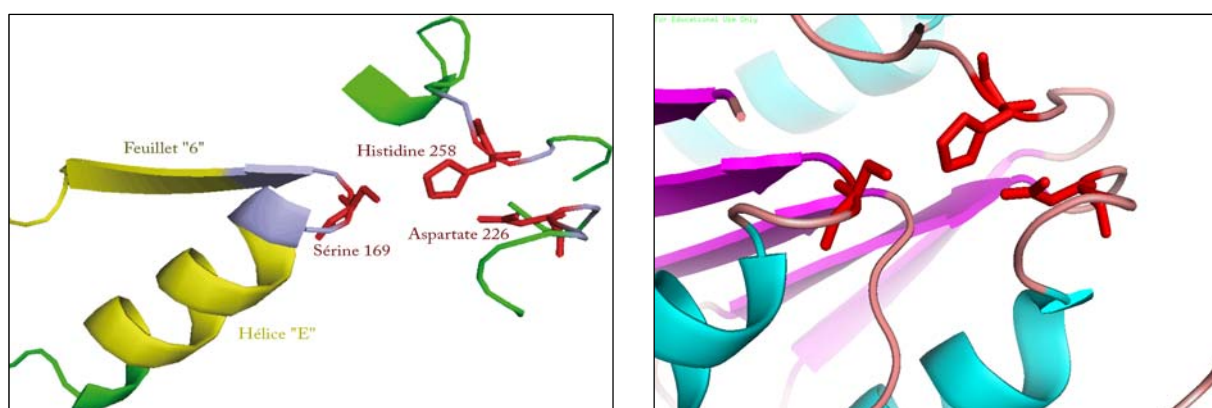


Figure 29: Organisation structurale de la triade catalytique d'*IgEst3* et de la carboxylestérase de *Francisella tularensis*.

Les structures secondaires sont schématisées selon la représentation standard. La forme développée illustre la position des trois acides aminés Ser, His et Asp intervenant dans la triade catalytique. A gauche, visualisation de la triade modélisée d'*IgEst3*. A droite, visualisation de la structure de la région homologue correspondant à la séquence [*FtCe1*].

2.1.4. Bilan des analyses *in silico* pour la séquence codante *IgEst3*

La première observation, est la probable absence de signaux peptidiques connus ainsi que l'absence probable de N- et O-glycosylations et de formation de ponts disulfure.

Les différentes combinaisons d'alignements (via l'algorithme BLASTP) ont, d'une part, permis de cibler deux fonctions enzymatiques putatives pour *IgEst3* : carboxylestérase ou thioestérase. D'autre part, elles ont permis d'observer que les séquences homologues (hors alignement avec des séquences issues d'algues et de plantes) sont essentiellement des séquences de microorganismes et plus particulièrement de bactéries. De plus, l'alignement avec les séquences de la banque PDB a permis d'identifier les positions putatives des acides aminés de la triade catalytique Ser¹⁶⁹, Asp²²⁶ et His²⁵⁸.

Les premières informations concernant la fonction enzymatique putative d'*IgEst3* avaient été obtenues par lecture des annotations du cluster ISL 00006895 disponibles dans la base de données TBestDB. La séquence du cluster avait présenté une homologie avec la famille de protéines PFAM PF02230, correspondant à une famille d'enzymes à activité phospholipase A2 et lysophospholipase indépendante du calcium. De plus, une homologie avec les séquences d'une lysophospholipase du riz [*OsLp2*] et d'une phospholipase/carboxylestérase d'*Arabidopsis thaliana* [*AtPLCe2*] avaient également été caractérisée. La séquence complète *IgEst3* a été comparée avec ces deux dernières séquences (Tableau 47). Les pourcentages d'identité sont proches de 30 % et les pourcentages de similarité sont proches de 40 %.

	[<i>OsLp2</i>]	[<i>AtPLCe2</i>]
<i>Score</i>	62,8	65,5
<i>Identité %</i>	32 %	29 %
<i>Similarité %</i>	43 %	40 %

Tableau 47: Pourcentages d'identité et de similarité des séquences ayant permis l'annotation du cluster ISL 00006895 dans la base de données TBestDB par rapport à la séquence protéique complète *IgEst3*.

Les valeurs obtenues correspondent à l'homologie de la séquence *IgEst3* par rapport aux séquences ayant permis l'annotation de son cluster. Ces valeurs sont à comparer aux valeurs obtenues lors de la recherche d'homologie réalisée au cours de cette thèse. Les séquences [*OsLp2*] et [*AtPLCe2*] n'apparaissent plus comme les plus homologues à la séquence *IgEst3*. Les pourcentages d'identité et de similarité des séquences homologues ont peu évolué. En revanche, les valeurs de score sont meilleures dans le cas des séquences homologues récemment identifiées.

L'utilisation du logiciel PipeAlign a permis de mettre en évidence que la séquence *IgEst3* est retrouvée dans le groupe où les séquences sont majoritairement de fonction inconnue. Cependant, pour ce même groupe, la fonction connue la plus représentée est la fonction thioestérase. Le second groupe n'incluant pas la séquence *IgEst3*, est majoritairement constitué de carboxylestérases.

L'étude *in silico* d'*IgEst3* a mis en évidence qu'une partie de la séquence en N-terminal de la protéine ne semble pas pouvoir se conformer (Figure 30). Cette partie, d'environ 40 acides aminés (à gauche), correspond à une zone non définie de la protéine n'appartenant pas à la structure identifiée comme étant celle d'une α/β hydrolase. En effet, la structure d'*IgEst3* est une structure modélisée à partir de structures connues α/β hydrolase. Tout ce qui n'appartient pas aux structures connues ne peut être modélisé et la conformation des parties C et N terminales n'a donc pas été calculée. Les structures connues d' α/β hydrolase ne présentent pas ce type de partie N-terminal et son rôle est donc inconnu

et amène à l'hypothèse de spécificités propres aux microalgues. Cette observation est à mettre en relation avec l'absence de signaux peptidiques identifiables dans la séquence traduite d'*IgEst3*.

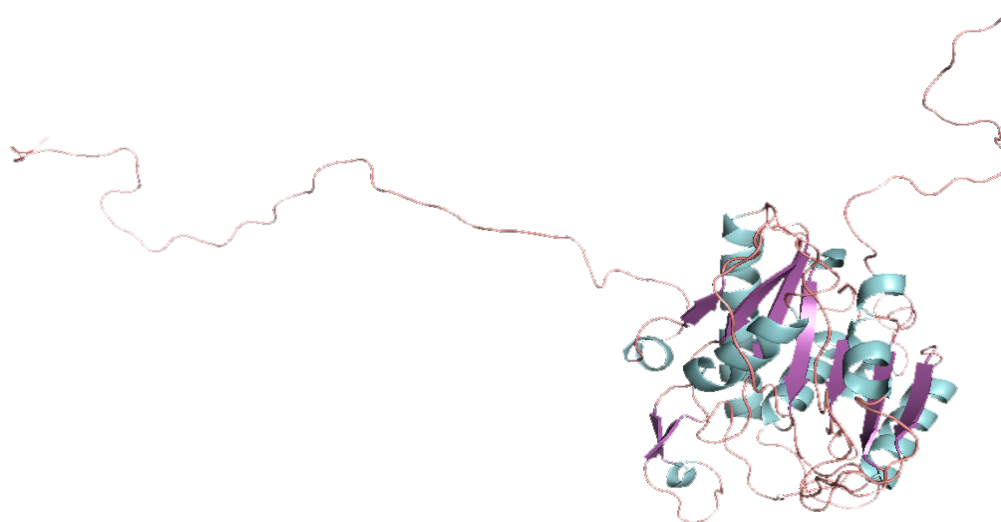


Figure 30: Modélisation tridimensionnelle complète d'*IgEst3*

2.2. Etude d'*IgTeCe*

La séquence codante *IgTeCe* est une séquence de 1146 pb codant une protéine putative de 381 acides aminés. La comparaison de la séquence génomique avec l'ADNc indique que le gène *IgTeCe* contient trois introns, le premier entre les positions 590 et 754, le second entre les positions 827 et 951 et enfin le troisième entre les positions 1049 et 1172. La figure 31 présente la séquence codante et la séquence protéique déduite d'*IgTeCe*. La séquence incomplète (cluster ISL00007519) issue de la base de données TBestDB est surlignée en bleu. Le pentapeptide GX SXG, indiqué en vert, est sous la forme GSSQG. Cette séquence, rare dans le cas des estérases, est à notre connaissance uniquement retrouvée pour certaines carboxylestérases de plantes supérieures (Souleyre *et al.*, 2011).

2.2.1. Prédictions des caractéristiques d'IgTeCe

```

-----|-----|-----|-----|-----|-----|
1 atggcgccgctcttctcgtcaaagccctccagaccggcgcccgagggccggtgactcgc 60
1 M G G R L R Q S P P D R R A Q G P V T R 20

-----|-----|-----|-----|-----|-----|
61 cgccgtttcttctcagccggcccccacatggacgtccggcggtccgccaagcagaccctt 120
21 R R F F S A G P H M D V R R S A K Q T L 40

-----|-----|-----|-----|-----|-----|
121 cttcagtccttcaatagcatgaagcgtcgacgcgagcggtccaacgtatccacatggag 180
41 L Q S F N S M K R R S D G V Q R I H M E 60

-----|-----|-----|-----|-----|-----|
181 gggctggcattcaaacgagcgtccggcctcgaggtattcgtcggacgcatccagaatc 240
61 G L A F K R A L R P R R Y S S D A S R I 80

-----|-----|-----|-----|-----|-----|
241 gccgcttctccaagtgcgaatcctcgtctcggacggcaagaggacgcgagaccgat 300
81 A A S S K S Q S S L S D G K E D S E T D 100

-----|-----|-----|-----|-----|-----|
301 ttggatgacgactgcagcgatgcctcgatgatcgacggatatctggcgccctcaggt 360
101 L D D D C S D A S M I D G Y L A A P S G 120

-----|-----|-----|-----|-----|-----|
361 cggattcggcgacgacgtccgaggatcccggtatcgtggtcatcgagccctcgtgcca 420
121 R I R R T T S E D P G I V V I E P S G P 140

-----|-----|-----|-----|-----|-----|
421 cacacgcacaccgccattctattgcacggcatgtactgctctccggagagctccgataca 480
141 H T H T A I L L H G M Y C S P E S S D T 160

-----|-----|-----|-----|-----|-----|
481 ttgtcggtctacgtgctgtcgaatcccttggcctgctgccagggattaaatgtgc 540
161 F V G L P A A V K S L G L L P G I K Y V 180

-----|-----|-----|-----|-----|-----|
541 ttcccgcatgcgcgcgtcggaccatcagctggccgacaggcccgaggcaaacgttgcc 600
181 F P H A P R R T I S W P T G P E A N V A 200

-----|-----|-----|-----|-----|-----|
601 tcttggtacaattactacacgcgcgggatggcgagctcgagcatgacgttctgaatgag 660
201 S W Y N Y Y T R R D G E L E H D V L N E 220

-----|-----|-----|-----|-----|-----|
661 ggcacacctggcttcgaaacccggcgatccactccattgtcgagcgcgaggtcgcgtt 720
221 A H L A S Q T R R I H S I V E R E A A L 240

-----|-----|-----|-----|-----|-----|
721 cttggggcgacgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 780
241 L G G D A R R I M L G C S S Q G G T V A 260

-----|-----|-----|-----|-----|-----|
781 ctccatgctacgatcaattaccatcgccctctcggcgactgctttgtcttcgctcatgt 840
261 L H A T I N Y H R P L G A L L C L R S C 280

-----|-----|-----|-----|-----|-----|
841 ctcatcgattcggctcagctttccgagagacaagcgtcccccgcgcgcgcgcgcgcgcgc 900
281 L I D S V T F P R D K R S P A A G T P V 300

-----|-----|-----|-----|-----|-----|
901 ttcgtcttcgcccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 960
301 F V F A A A Q D K V Y A P Q L Q Y R G F 320

-----|-----|-----|-----|-----|-----|
961 tctctgctggcgactctggcttcacgtggagtggcatgtggaacctcatctgacgcac 1020
321 S L L A D S G F H V E W H V E P H L T H 340

-----|-----|-----|-----|-----|-----|
1021 tgggacgagtcgcgcacagagaaacgcgtagtcgctgcatggatcgcgcgcacagtcgca 1080
341 W D E S R T E K R V V A A W I A R T V R 360

-----|-----|-----|-----|-----|-----|
1081 gagcgcgggctccgagagcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 1140
361 E R G L R E R E R S V S P E L I L E P L 380

-----
1141 ttctag 1146
381 F * 381

```

Figure 31: Séquences codante et protéique d'IgTeCe.

Le pentapeptide correspondant à la séquence consensus GX SXG est surligné en vert et composé comme suit : GSSQG. La séquence correspondant à la séquence ISL0007519 issue de la base de donnée EST apparait en bleu.

Recherche d'un signal peptidique

La valeur des trois signaux C, S et Y n'est pas significative et indique l'absence de signaux peptidiques détectés pour la séquence *IgTeCe*. Le graphique obtenu par SignalP est présenté en annexe 6.

Evaluation de la masse moléculaire et du point isoélectrique

La séquence codante *IgTeCe* code une protéine de masse moléculaire théorique de 42,31 kDa et de pI théorique de 9,8.

Prédiction de la formation de ponts disulfures

La séquence d'*IgTeCe* contient 4 cystéines indiquées en rouge sur fond gris en figure 31. Le logiciel de prédiction indique que la formation de ponts disulfures est peu probable.

Prédiction de la présence de sites de glycosylation

Les résultats obtenus indiquent l'absence de signaux peptidiques et de sites consensus, rendant improbable la formation de N-glycosylations. Le tableau des positions des acides aminés asparagines dans *IgTeCe* est présenté en annexe 7.

Le résultat de la recherche de sites d'O-glycosylation pour *IgTeCe* a mis en évidence 34 résidus compatibles. Cependant, l'absence de signaux peptidiques reconnaissables rend la modification par O-glycosylation peu probable. Les positions d'O-glycosylation détectées par chaque algorithme sont présentées en annexe 8.

2.2.2. Prédictions de la fonction enzymatique par recherche d'homologies

Recherche par l'algorithme BLASTP

Les alignements retenus pour prédire la fonction d'*IgTeCe* sont présentés dans le tableau 48.

Numéro d'alignement	Méthodologie choisie	Banque de données choisie	Organismes choisis	Protéines putatives incluses
1	BLASTP	NR	Tous	Oui
2	BLASTP	RefSeq	Tous	Non
3	BLASTP	RefSeq	Algues	Non
4	BLASTP	SwissProt	Tous	Non
5	BLASTP	PDB	Tous	Non
6	DELTA-BLAST	NR	Tous	Non
7	DELTA-BLAST	NR	Plantes	Non
8	DELTA-BLAST	NR	Algues	Non
9	PHI-BLAST	NR	Tous	Non

Tableau 48: Liste des alignements retenus pour la protéine *IgTeCe*

- Alignement 1 : BLASTP/NR

Cet alignement est réalisé par l'algorithme BLASTP en utilisant la banque NR. Les quatre séquences homologues présentant les plus hauts scores sont d'origine marine (Tableau 49). Les séquences

$|EhHp2|$ et $|EhHp3|$ sont des séquences de protéines hypothétiques de la microalgue *Emiliana huxleyi*. La séquence $|AaHp1|$ est une séquence de protéine hypothétique de la microalgue *Aureococcus anophagefferens*. La séquence homologue $|GnPlCe1|$ est une séquence de bactérie marine et présente une fonction putative de phospholipase ou carboxylestérase. Dans le cas de cette séquence, le score de l'alignement, les pourcentages d'identité et de similarité sont inférieurs à ceux obtenus pour les séquences homologues de microalgues.

	$ EhHp2 $	$ EhHp3 $	$ AaHp1 $	$ GnPlCe1 $
Score	174	125	127	94,7
Identité %	47 %	35 %	42 %	32 %
Similarité %	55 %	50 %	57 %	51 %

Tableau 49: BLASTP d'*IgTeCe* contre les séquences des banques NR.

- Alignement 2 : BLASTP/RefSeq

Cet alignement, réalisé avec les séquences référencées dans RefSeq (Tableau 50), a permis d'identifier quatre séquences homologues issues de microorganismes marins proches. La première séquence $|GnPlCe1|$ est une séquence putative de carboxylestérase ou de phospholipase issue de la bactérie *Glaciecola nitratreducens*. Il s'agit de la même séquence que celle évoquée dans l'alignement 1 BLASTP/NR. Des homologues de séquence sont ensuite retrouvées avec des séquences putatives de thioestérases issues des bactéries *Glaciecola pallidula* $|GpTe1|$, *Glaciecola chatamensis* $|GcTe1|$ et *Glaciecola agarilytica* $|GaTe1|$. Il est intéressant de noter que la séquence $|GcTe1|$ est également homologue à *IgEst3*.

	$ GnPlCe1 $	$ GpTe1 $	$ GcTe1 $	$ GaTe1 $
Score	94,7	90,1	85,1	84,7
Identité %	32 %	31 %	27 %	27 %
Similarité %	51 %	50 %	46 %	46 %

Tableau 50: BLASTP d'*IgTeCe* contre les séquences de la banque RefSeq.

- Alignement 3 : BLASTP/RefSeq/Algues

Cet alignement, restreint aux séquences issues d'algues, fait apparaître deux groupes (Tableau 51). Le premier contient les enzymes issues d'algues rouges. La séquence $|CmLp2|$ est une lysophospholipase putative issue de *Cyanidioschyzon merolae* et la séquence $|GsLp2|$ est une lysophospholipase putative issue *Galdiera sulphuraria*. Le second groupe contient des enzymes issues de microalgues. La séquence $|CsPlCe1|$ est une séquence putative pour deux fonctions carboxylestérase ou phospholipase, issue de *Coccomyxa subellipsoidea* (chlorophycée) et la séquence $|NgLp2|$ a déjà été précédemment décrite. Cet alignement fait apparaître la fonction lysophospholipase. Toutefois, il est important de

remarquer que les valeurs de score sont proches de 50. Par comparaison avec les valeurs observées lors de l'alignement précédent ces valeurs ont été nettement diminuées.

	<i> CmLp2 </i>	<i> GsLp2 </i>	<i> CsPlCe1 </i>	<i> NgLp2 </i>
Score	50,8	50,4	49,7	48,1
Identité %	31 %	26 %	28 %	28 %
Similarité %	42 %	38 %	39 %	41 %

Tableau 51: BLASTP d'*IgTeCe* contre les séquences de la banque RefSeq issues d'algues (macro- ou micro-).

- Alignement 4 : BLASTP/SwissProt

Comparée aux séquences de la base de données SwissProt, la séquence *IgTeCe* présente de nouveau des homologues avec des séquences putatives de thioestérases (*|ScpTe1|*, *|GzTe1|*, *|DdTe1h2|*). La séquence *|HsLYPLIS|*, déjà homologue à la séquence d'*IgEst3*, correspond à la lysophospholipase 1 humaine (Tableau 52).

	<i> ScpTe1 </i>	<i> GzTe1 </i>	<i> DdTe1h2 </i>	<i> HsLYPLIS </i>
Score	68,9	60,8	60,1	59,7
Identité %	28 %	28 %	29 %	24 %
Similarité %	47 %	46 %	45 %	43 %

Tableau 52: BLASTP d'*IgTeCe* contre les séquences de la banque Swissprot.

- Alignement 5 BLASTP/PDB

Le dernier alignement BLASTP est réalisé contre la banque de données PDB de manière à définir la séquence consensus et les acides aminés composant la triade catalytique (Tableau 53). Les séquences homologues obtenues par cet alignement sont à séparer en deux groupes. Le premier regroupe des séquences issues d'organismes procaryotes : *|PaCe1|* la séquence de carboxylestérase issue de la bactérie *Pseudomonas aeruginosa* et les séquences *|PfCe2S|* et *|FrCe1|* précédemment décrites. Le second groupe contient les séquences de lysophospholipase humaine *|HsLYPLIS|* et de thioestérase humaine *|HsTe1|* qui ont également été décrites précédemment. L'ensemble de ces séquences a permis de localiser la position des acides aminés susceptibles de former la triade catalytique d'*IgTeCe*, la composition des régions bordant ces acides aminés reste peu conservée (Figure 32).

	<i> PaCe1 </i>	<i> HsLYPLIS </i>	<i> PfCe2S </i>	<i> FrCe1 </i>	<i> HsTe1 </i>
Score	61,2	60,5	58,2	54,7	49,3
Identité %	31 %	25 %	27 %	24 %	29 %
Similarité %	41 %	43 %	40 %	44 %	48 %

Tableau 53: BLASTP d'*IgTeCe* contre les séquences de la banque PDB.

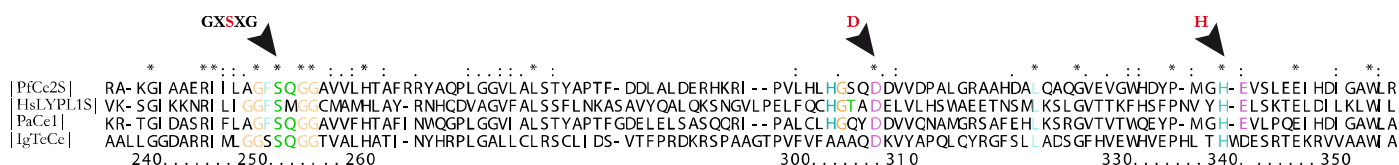


Figure 32: Alignement des séquences homologues obtenues dans la banque de données PDB. Les acides aminés signalés en rouge correspondent aux résidus de la triade catalytique S, D, H d'IgTeCe.

- Alignement 6 DELTA-BLAST/NR

L'algorithme BLASTP a permis d'identifier trois fonctions putatives d'IgTeCe : carboxylestérase, thioestérase ou lysophospholipase. Pour affiner ces prédictions de fonction et les hiérarchiser, des alignements complémentaires ont été réalisés avec DELTA-BLAST (Tableau 54). Il est intéressant de noter les différentes « origines » des séquences homologues à IgTeCe. Ce sont toutes des séquences issues d'eucaryotes et elles codent toutes pour des thioestérases. La séquence |CeATH1| est une séquence putative issue de *Caenorhabditis elegans*. La séquence |BmTe2| est une séquence issue du yak (*Bos mutus*). La séquence |RnTe2| est issue de *Rattus norvegicus* et la séquence |MmTe2| est issue de *Mus musculus*. L'ensemble des scores obtenus est supérieur à 150 mais les pourcentages d'identité observés sont faibles, ils sont inférieurs par exemple à ceux observés en utilisant l'algorithme BLASTP.

	CeATH1	BmTe2	RnTe2	MmTe2
Score	159	157	157	157
Identité %	24 %	22 %	22 %	22 %
Similarité %	40 %	35 %	35 %	35 %

Tableau 54: DELTA-BLAST d'IgTeCe contre les séquences des banques NR.

- Alignement 7 DELTA-BLAST/NR/Plantes

L'alignement DELTA-BLAST a également été réalisé spécifiquement pour les séquences issues de plantes. Les scores d'alignement des séquences homologues sont très proches (Tableau 55). La séquence |TcEst1| est une séquence putative d'ester hydrolase issue de *Theobroma cacao*. Les deux séquences suivantes |RcTe1| et |RcTe2| sont des séquences putatives de thioestérases issues du ricin. |RcTe1| et |JcPlCe1| sont des séquences également homologues à IgEst3.

	TcEst1	RcTe1	RcTe2	JcPlCe1
Score	123	121	120	120
Identité %	20 %	20 %	19 %	21 %
Similarité %	33 %	34 %	33 %	33 %

Tableau 55: DELTA-BLAST d'IgTeCe contre les séquences des banques NR issues de plantes

- Alignement 8 DELTA-BLAST/NR/Algues

Le DELTA-BLAST suivant a été réalisé par comparaison aux séquences d'algues présentes dans la banque de données NR (Tableau 56). Les deux premières séquences homologues sont des séquences putatives de lysophospholipase issues de la macroalgue *Ectocarpus siliculosus* [*EsLp2*] et de la microalgue *Nannochloropsis gaditana* [*NgLp1*]. La troisième séquence [*CsPlCe1*] est une séquence putative pour deux fonctions, carboxylestérase ou phospholipase de la microalgue *Coccomyxa subellipsoidea*.

	[<i>EsLp2</i>]	[<i>NgLp1</i>]	[<i>CsPlCe1</i>]	[<i>TpPl1</i>]
Score	129	125	120	118
Identité %	22 %	24 %	23 %	25 %
Similarité %	39 %	36 %	39 %	37 %

Tableau 56: DELTA-BLAST d'*IgTeCe* contre les séquences des banques NR issues d'algues.

- Alignement 9 PHI-BLAST/NR

En complément de ces analyses, l'algorithme PHI-BLAST a été utilisé (Tableau 57). En tenant compte de la présence du pentapeptide consensus GX SXG, les meilleurs scores obtenus sont pour les séquences putatives décrites à l'alignement 2 BLASTP/RefSeq, [*GnPlCe1*], [*GpTe1*] et [*GcTe1*]. La séquence [*MgCe1*] est une séquence putative de carboxylestérase issue de la bactérie *Methylovorus glucosetrophus*. Cet algorithme n'a pas permis d'affiner la sélection des fonctions enzymatiques possibles pour *IgTeCe*.

	[<i>GnPlCe1</i>]	[<i>GpTe1</i>]	[<i>MgCe1</i>]	[<i>GcTe1</i>]
Score	77,7	71,6	70,8	69,7
Identité %	32 %	31 %	32 %	27 %
Similarité %	51 %	51 %	47 %	46 %

Tableau 57: PHI-BLAST d'*IgTeCe* contre les séquences des banques NR sans restriction d'organismes, l'alignement tient compte de la présence du pentapeptide GX SXG.

- Recherche de séquences homologues dans les bases de données ESTHER, MELDB et LED

Il apparaît, au vu des précédents alignements, une large variété de fonctions putatives pour la protéine *IgTeCe*. Une recherche de séquences homologues dans les bases de données ESTHER et LED a été entreprise. La séquence [*SaCe1*] caractérise la meilleure homologie de cette recherche. Il s'agit d'une séquence putative de carboxylestérase, issue de la bactérie *Shewanella amazonensis*. L'alignement contre la banque de données MELDB identifie la séquence [*CbPlCe1*], caractérisée comme une séquence putative des fonctions carboxylestérase ou phospholipase, issue de la bactérie *Coxiella burnetii*. En conclusion, la recherche de séquences dans ces bases de données spécifiques

prédit une homologie à des séquences procaryotes, sans permettre de distinguer une fonction enzymatique précise.

Recherche avec le logiciel PipeAlign

La séquence *IgTeCe* a également été alignée avec PipeAlign (Figure 33). A l'issue de l'ensemble des traitements, il est observé que le groupe A majoritairement composé de carboxylestérases n'inclut pas la séquence *IgTeCe*. Le groupe B, incluant la séquence *IgTeCe*, est composé quasi-unanimement de protéines hypothétiques et de protéines de fonction inconnue. Le résultat de PipeAlign ne permet donc pas de qualifier plus précisément la fonction de la protéine *IgTeCe* et met en évidence la difficulté à identifier sa fonction *in silico*.

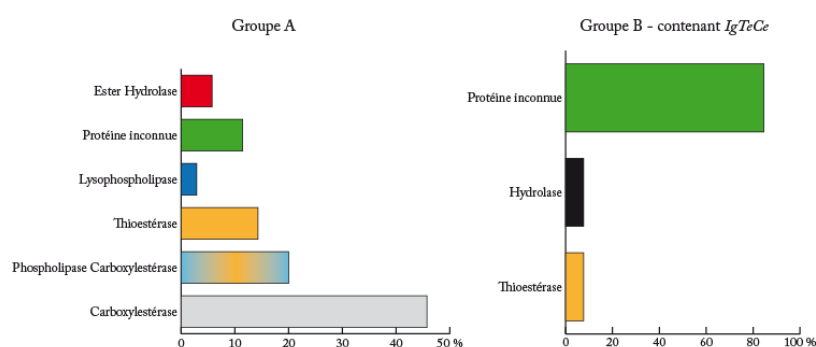


Figure 33: Composition des deux groupes d'enzymes obtenus par le logiciel PipeAlign

2.2.3. Prédictions de l'organisation structurale

La prédiction de la structure secondaire pour la protéine *IgTeCe* a été réalisée avec l'algorithme YASPIN (Figure 34 A) d'une part et déduit à partir de la structure tridimensionnelle (Figure 34 B) d'autre part.

Dans les deux cas, une succession d'hélices α et de feuillets β , permettant de placer la protéine dans la famille des hydrolases α/β , est mise en évidence. Le pentapeptide GX SXG est placé, à la jonction entre une structure en feuillet et une structure en hélice laissant la sérine de la triade catalytique accessible entre ces deux structures.

Les principales différences de prédiction entre les deux analyses concernent les parties N-terminal et C-terminal. En effet, l'algorithme YASPIN détecte 6 hélices en N-terminal (120 premiers acides aminés) alors qu'elles n'apparaissent pas dans la prédiction de la structure tridimensionnelle (Figure 34 B). Les prédictions de structure secondaire pour le reste de la protéine sont identiques mis à part le double feuillet « 4 et 5 », non détecté par l'algorithme YASPIN. De plus, deux feuillets sont absents en position C-terminal, de la prédiction déduite de la structure tridimensionnelle mais présents dans le modèle proposé par YASPIN.

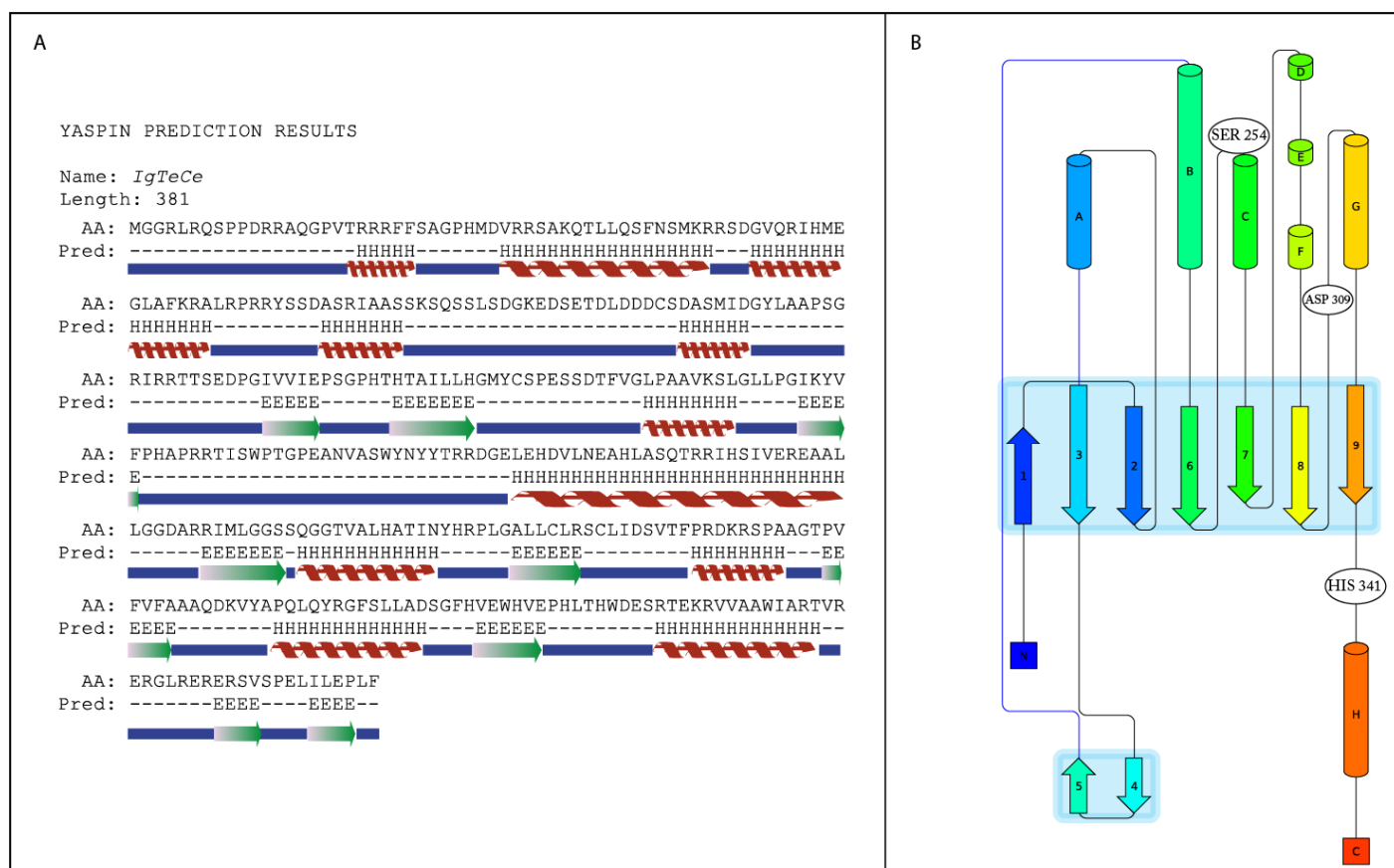


Figure 34: Prédiction de la structure secondaire d'IgTeCe

A - Schématisation de la structure secondaire de la protéine *IgTeCe* après analyse avec l'algorithme YASPIN. B - Illustration de la structure secondaire issue de la structure tridimensionnelle modélisée.

La modélisation tridimensionnelle de la protéine *IgTeCe* par le logiciel Modeller a été réalisée à partir des séquences de référence résultant de l'alignement 5 BLASTP/PDB. Pour la modélisation d'*IgTeCe*, trois séquences ont été conservées : *[HsLYPL1S]*, *[PfCe2S]* et *[FtCe1]*. La séquence bénéficiant du meilleur score *[PaCe1]* n'a pas pu être utilisée par l'algorithme. La figure 35 illustre la conformation structurale calculée pour la triade catalytique. Le calcul du positionnement des composés de la triade catalytique semble cohérent par rapport à la structure de la lysophospholipase humaine.

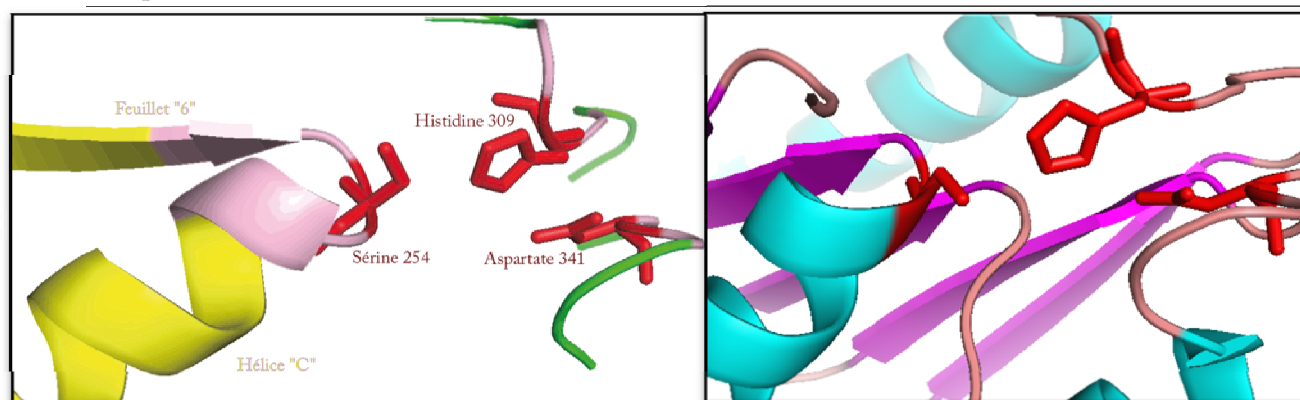


Figure 35: Organisation structurale de la triade catalytique d'*IgTeCe* et de la lysophospholipase humaine [*HsLYPLIS*].

A gauche, schématisation de la structure tridimensionnelle d'*IgTeCe*. La forme développée illustre la position des trois acides aminés Ser, His et Asp intervenant dans la triade catalytique. A droite, illustration de la région homologue correspondant à la séquence alignée [*HsLYPLIS*].

2.2.4. Bilan des analyses *in silico* pour la séquence codante *IgTeCe*

L'analyse de la séquence protéique d'*IgTeCe* a montré la probable absence de signaux peptidiques connus ainsi que l'absence probable de N- et O-glycosylations et de formation de ponts disulfure.

La stratégie via l'algorithme BLASTP n'a pas permis de cibler une fonction enzymatique précise. Les fonctions lysophospholipase, carboxylestérase, phospholipase et thioestérase sont toutes les quatre mises en évidence sans qu'aucune ne soit prédominante.

Une partie de la séquence codante d'*IgTeCe* est présente dans la base de données TBestDB. Il s'agit du cluster ISL00007519. Les annotations indiquent qu'une homologie avec le motif PFAM PF02230 avait été identifiée. Une homologie avec les séquences d'une lysophospholipase du champignon *Neurospora crassa* [*NcLp1*] et d'une carboxylestérase putative de la bactérie marine *Chromohalobacter salexigens* [*CsCe1*], résistante aux environnements à salinité élevée, avait également été mis en évidence.

Au cours de cette étude, l'alignement BLASTP a été réalisé entre ces deux dernières séquences et la séquence complète d'*IgTeCe*. Les pourcentages de similarité, obtenus par alignement local BLASTP, sont proches de 45 % et les pourcentages d'identité proches de 30 % (Tableau 58).

	[<i>NcLp1</i>]	[<i>CsCe1</i>]
Score	55,5	76,6
Identité %	29 %	31 %
Similarité %	46 %	47 %

Tableau 58: Pourcentage d'identité et de similarité des séquences ayant permis l'annotation du cluster ISL00007519 dans la base de données TBestDB par rapport à la séquence protéique complète *IgTeCe*.

Les valeurs obtenues ici par BLASTP correspondent à la qualité de l'alignement entre la séquence complète d'*IgTeCe* et les séquences [*NcLp1*] et [*CsCe1*]. Par comparaison avec les valeurs obtenues

lors de la recherche d'homologie réalisée au cours de cette thèse, les pourcentages d'identité et de similarité sont proches. Néanmoins, il faut noter deux paramètres importants, les valeurs de scores obtenues au cours de cette étude sont meilleures et une fonction enzymatique supplémentaire est apparue, la fonction thioestérase.

L'alignement avec les séquences de la banque PDB a permis d'identifier les positions putatives des acides aminés de la triade catalytique Ser²⁵⁴, Asp³⁰⁹ et His³⁴¹.

La seconde stratégie, utilisant le logiciel Pipealign, met en évidence l'appartenance de la séquence *IgTeCe* au groupe dont les séquences sont majoritairement de fonctions inconnues.

La démarche de prédiction *in silico* de l'activité enzymatique pour *IgTeCe* n'est donc pas concluante. Cependant, il faut savoir que l'analyse *in silico* de la séquence *IgTeCe* a évolué au cours du temps. En effet, le cluster ISL00007519, identifiant la séquence d'*IgTeCe*, a fait l'objet d'une prédiction de fonction en 2005. Par la suite, après l'obtention de l'intégralité de la séquence codante d'*IgTeCe* au laboratoire, le résultat de cet alignement est resté identique jusqu'en 2010. A partir de la fin 2010, de nouveaux alignements réalisés à l'aide des différents algorithmes ont commencé à inclure également la fonction thioestérase. Dans certains cas, la fonction thioestérase est même apparue comme prédominante. L'évolution du résultat d'alignement de la séquence *IgTeCe* est un réel changement de sous-classe enzymatique : E.C. 3.1.1.x, vers E.C. 3.1.2.

Historiquement, la prédiction de la fonction enzymatique d'*IgTeCe* est « carboxylestérase ». En effet, la conformation générale de la protéine par modélisation tridimensionnelle est particulièrement homologue à des α/β hydrolases de type « phospholipase » ou « carboxylestérase » (Figure 36).

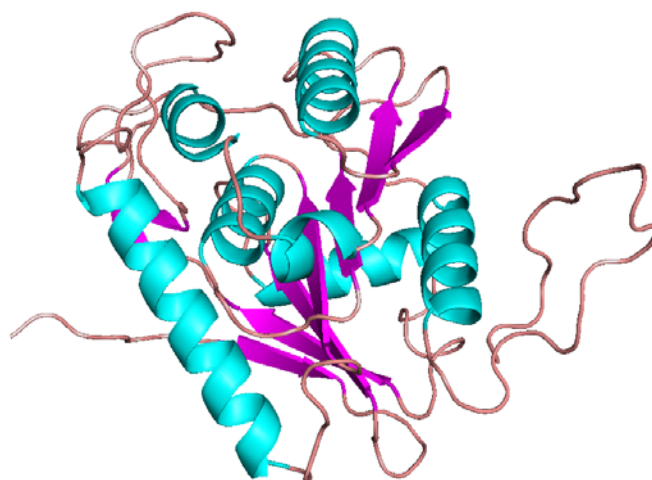


Figure 36: Prédiction de la structure de la protéine *IgTeCe*.
120 acides aminés du domaine N-terminal (à gauche) ne sont pas visibles sur cette représentation

De plus, la présence du pentapeptide GX SXG confirme cette prédiction. Seule une faible proportion des enzymes de fonction thioestérase adoptent la conformation des α/β hydrolases. Au regard des informations accumulées, seule la caractérisation biochimique *in vivo* va permettre de déterminer la fonction enzymatique.

L'analyse *in silico* révèle une particularité structurale. Une partie importante de la séquence d'*IgTeCe* en N-terminal de la protéine ne semble pas pouvoir être modélisée (Figure 37). En absence de modélisation, cela signifie que les structures de références utilisées ne présentent pas de domaines N-terminal correspondant. La longueur significative (environ 125 acides aminés) de cette partie N-terminal correspond à un domaine inconnu de la protéine, non annoté lors des alignements de séquences et représentant quasiment un tiers de la protéine totale.

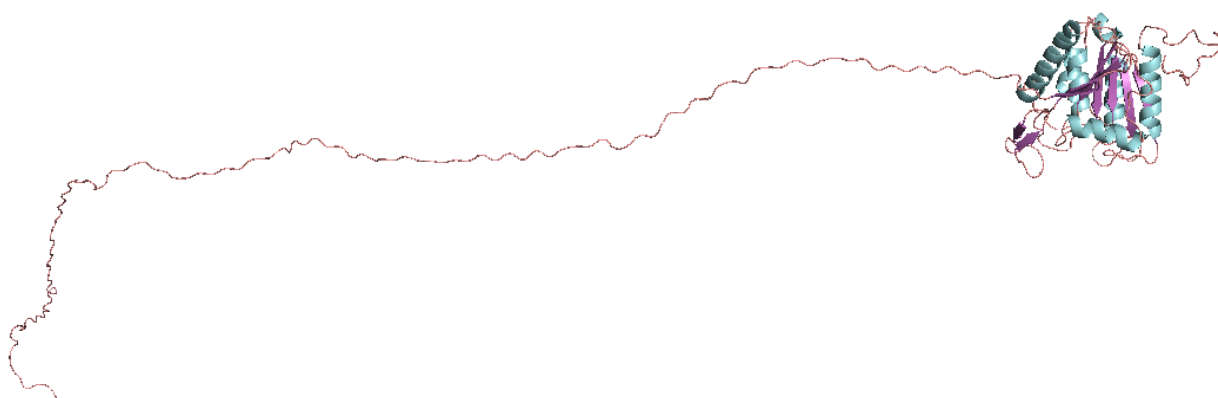


Figure 37: Modélisation « globale » de l'ensemble de la protéine *IgTeCe*

3. Expression hétérologue

3.1. Préambule

Les travaux de clonage pour l'expression des deux séquences codantes *IgEst3* et *IgTeCe* dans la levure ont été effectués au laboratoire avant le début de ce travail de thèse. Les séquences ont été clonées sous forme complètes. Une alanine en position 2 avait été ajoutée aux deux protéines, dans le but de favoriser l'obtention de la protéine d'intérêt. Le contrôle « témoin LacZ » est une construction plasmidique *pYES2.1-V5-His-LacZ* fournie par Invitrogen™ permettant d'exprimer la β -galactosidase. Cette protéine témoin est complétée en C-terminal d'une étiquette poly-histidine. L'obtention de la β -galactosidase sert d'indicateur pour la validation des conditions d'expression dans la levure.

Les travaux de clonage pour l'expression dans la bactérie ont été réalisés au cours de ce travail de thèse et uniquement pour *IgTeCe*. La séquence codante d'*IgTeCe* a été clonée sous forme native.

3.2. Expression dans la levure *Saccharomyces cerevisiae*

3.2.1. Validité des systèmes d'expression contenant *pYE-IgTeCe* et *pYE-IgEst3*

Avant de débiter les travaux de cette thèse, un premier séquençage a été réalisé sur ces plasmides après clonage et extraction. Celui-ci a permis de confirmer la présence des séquences codantes *IgEst3* (taille totale du plasmide : 6882 pb) et *IgTeCe* (taille totale du plasmide : 7035 pb) au sein des deux constructions.

Lors de ce travail, une vérification par PCR de la présence des inserts dans chaque système d'expression a été entreprise après transformation dans la souche d'expression INVSc1.

Extraction des plasmides

Afin de vérifier l'extraction des plasmides à partir de la levure, une analyse sur gel d'agarose a été effectuée. La figure 38 montre la présence des trois ADN plasmidiques. Les deux bandes observées pour chaque ADN correspondent à la forme relâchée (bande la plus haute entre 10000 et 21000 pb) et à la forme surenroulée du plasmide (bande inférieure à environ 5000 pb). Ces résultats indiquent une bonne extraction des plasmides depuis la levure.

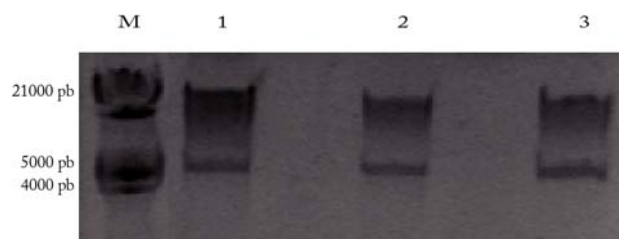


Figure 38: Electrophorèse sur gel d'agarose d'ADN plasmidiques issus de la levure.
M : Marqueur de taille, Piste 1: *pYES2.1-V5-His-LacZ*, Piste 2: *pYE-IgEst3*, Piste 3: *pYE-IgTeCe*.

Analyse des plasmides

Les extraits plasmidiques ont ensuite été amplifiés par PCR. L'objectif est, d'une part, de confirmer que l'insert (*IgTeCe* ou *IgEst3*) est toujours présent après transformation et d'autre part, d'obtenir un fragment amplifié pouvant être analysé par séquençage.

Deux analyses ont été effectuées. La première PCR a été réalisée à l'aide d'amorces spécifiques des régions bordant le site de clonage, K7EXTRACTFW et K7EXTRACTRV. La figure 39 présente un focus de la séquence du site d'insertion dans le plasmide *pYES2.1*. Le site d'insertion est indiqué en rouge et les sites d'hybridation des amorces sont indiqués en bleu. La taille du fragment amplifié sera supérieure à la taille de l'insert puisque les amorces « K7EXTRACT » s'hybrident respectivement en amont et aval du site d'insertion (soit 234 bases en 5' et 183 bases en 3').



Figure 39: Séquence partielle du plasmide *pYES2.1*
Présentation des sites d'hybridation « K7EXTRACT ». Les nombres représentent les positions des nucléotides sur le plasmide.

La taille des fragments amplifiés présentés en figure 40 correspond à celle attendue pour chacun des deux systèmes d'expression : 1407 pb en présence de la séquence codante *IgEst3* (piste 1) et 1563 pb en présence de la séquence codante *IgTeCe* (piste 2). Les inserts sont donc bien présents dans les deux systèmes d'expression.

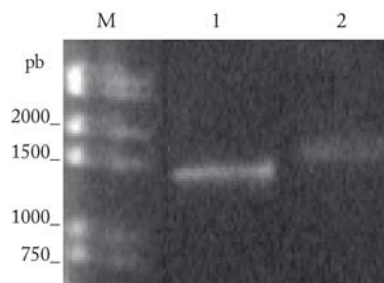


Figure 40: Electrophorèse sur gel d'agarose des fragments amplifiés par PCR avec les amorces « K7EXTRACT ». Vérification de la présence des inserts dans *pYES2.1*.

M : Marqueur de taille, Piste 1 : PCR depuis *pYE-IgTeCe*, Piste 2 : PCR depuis *pYE-IgEst3*.

La seconde analyse a été effectuée par amplification spécifique des séquences codantes *IgTeCe* et *IgEst3*. L'analyse des produits PCR par électrophorèse, a permis de confirmer la présence et d'identifier les séquences codantes d'*IgEst3* (990 pb, en piste 1) et d'*IgTeCe* (1143 pb, en piste 2) aux tailles attendues (Figure 41).

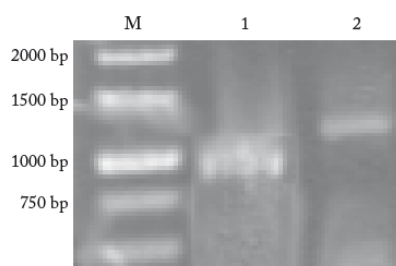


Figure 41: Electrophorèse sur gel d'agarose des fragments amplifiés spécifiques de chaque gène *IgEst3* et *IgTeCe*. M : marqueur de taille, Piste 1 : 6895-ATG2 x 6895-STOP2E, Piste 2 : 7519-ATG2 x 7519-STOP2E

3.2.2. Caractéristiques prédites des protéines recombinantes *IgEst 3* et *IgTeCe*

Pour faciliter la détection et la purification des protéines recombinantes, le vecteur *pYES2.1* a été utilisé pour ajouter un épitope V5 et un domaine poly-histidine (6x His) en position C-terminal. Cet ajout modifie les propriétés des protéines recombinantes par rapport aux protéines natives (Tableau 59). Il influe sur la longueur de la protéine, sur sa masse moléculaire et sur le point isoélectrique.

Nom de la protéine	Propriétés de la protéine native	Propriétés de la protéine recombinante
<i>IgEst3</i>	329 aa 35,41 kDa pI 8,66	363 aa 39 kDa pI 8,6
<i>IgTeCe</i>	381 aa 42,31 kDa pI 9,8	415 aa 46,13 kDa pI 9,55

Tableau 59 : Caractéristiques des protéines natives et recombinantes.

3.2.3. Expression et extraction de la protéine *IgEst3*

L'expression de la protéine *IgEst3* a été réalisée en parallèle de l'expression de la β -galactosidase, qui a servi de contrôle des conditions d'expression. Chaque étape d'extraction et de purification ont été effectuées selon le même protocole. Une fois la β -galactosidase produite, celle-ci disposant d'une étiquette poly-histidine, a servi de contrôle positif. L'extraction des protéines a été réalisée dans un tampon 50 mM phosphate de sodium, pH 8.

L'extraction et la purification par chromatographie d'affinité ont été évaluées sur SDS-PAGE (Figure 42).

Les résultats obtenus pour la β -galactosidase sont présentés en figure 42 A. La β -galactosidase (120 kDa) est clairement visible dans l'extrait brut du témoin LacZ induit (piste 1). Elle est également visible dans les fractions purifiées aux pistes 6 et 8. Comme attendu, les pistes adjacentes, correspondant aux fractions du témoin LacZ non induit, ne présentent pas de bandes au poids moléculaire de la β -galactosidase.

La figure 42 B présente les résultats obtenus pour *IgEst3* (induit et non induit). Aucune bande significative n'est observée aux pistes 5 et 7 au poids moléculaire attendu. Des bandes sont bien présentes mais elles sont également présentes pour les échantillons non induits. De plus, le poids moléculaire ne correspond pas. Une bande majeure est présente dans la fraction correspondant au lavage de la résine (Piste 9) mais elle observée entre 25 et 30 kDa alors que le poids moléculaire attendu est de 39 kDa.

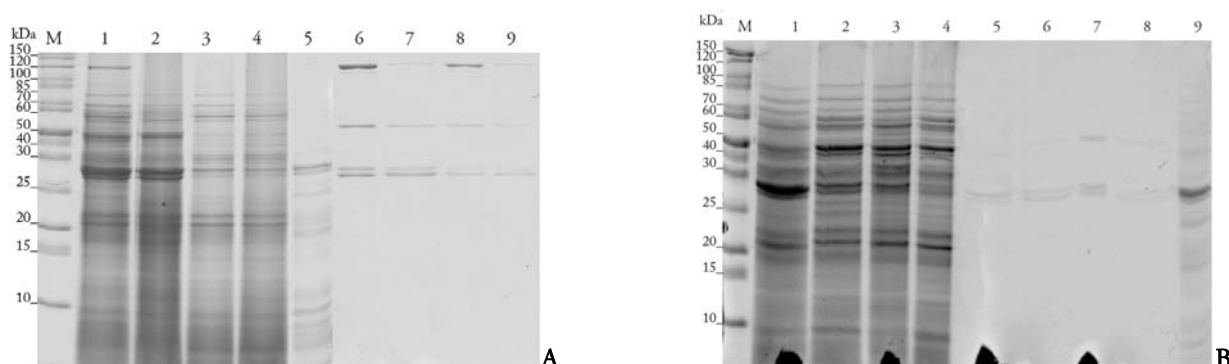


Figure 42: Electrophorèse sur SDS-PAGE après chromatographie d'affinité du témoin LacZ (A) et d'*IgEst3* (B).
 (A) M : Marqueur de poids moléculaire, Piste 1 : Extrait brut induit, Piste 2 : Extrait brut non induit, Piste 3 : Fraction non retenue (induit), Piste 4 : Fraction non retenue (non induit), Piste 5 : Lavage de la résine (induit), Piste 6 : Première élution (induit), Piste 7 : Première élution (non induit), Piste 8 : Seconde élution (induit), Piste 9 : Seconde élution (non induit).
 (B) M : Marqueur de poids moléculaire, Piste 1 : Extrait brut induit, Piste 2 : Extrait brut non induit, Piste 3 : Fraction non retenue (induit), Piste 4 : Fraction non retenue (non induit), Piste 5 : Première élution (induit), Piste 6 : Première élution (non induit), Piste 7 : Seconde élution (induit), Piste 8 : Seconde élution (non induit), Piste 9 : Lavage de la résine (induit).

La chromatographie d'affinité sur la résine « NiNTA » est fonctionnelle avec la β -galactosidase disposant de la même étiquette qu'*IgEst3* et exprimée dans les mêmes conditions.

Les analyses par Dot blot et Western blot, en utilisant un anticorps spécifique de l'étiquette poly-histidine, n'ont pas révélé de signaux pour *IgEst3* mais ont fonctionné pour la détection de la β -galactosidase.

3.2.4. Recherche d'activités lipolytiques pour *IgEst3*

En parallèle de l'expression et de l'extraction de la protéine *IgEst3*, des tests enzymatiques ont été réalisés avec les extraits protéiques par colorimétrie sur des substrats synthétiques : le para-nitrophényle caprylate (C8) et le para-nitrophényle palmitate (C16). Chaque test a comparé l'activité d'hydrolyse d'un échantillon induit par rapport à un échantillon non induit. Une activité plus importante est attendue lorsque l'échantillon est induit par rapport au non induit ou au témoin produisant la β -galactosidase.

Le résultat de ce test est négatif. Aucune différence d'activité enzymatique n'a été observée entre les différents échantillons. Ce résultat va dans le sens de l'absence de la protéine.

Différentes hypothèses de dysfonctionnement peuvent être écartées. Les conditions de culture, la composition du milieu, la source de carbone ou l'inducteur semblent en effet correctes et compatibles avec le système d'expression utilisé car l'obtention de la β -galactosidase est constatée.

En conclusion, l'ensemble des expérimentations réalisées pour *IgEst3* n'ont pas permis de caractériser de protéines correspondant à la protéine *IgEst3*. Elles n'ont pas non plus permis de démontrer le bon fonctionnement de la construction *pYE-IgEst3*. L'absence d'activité lipolytique détectable indique en plus que dans le cas d'une très faible production de la protéine *IgEst3*, celle-ci semble inactive. L'hypothèse la plus probable est l'absence de la protéine *IgEst3*.

3.2.5. Expression et extraction de la protéine *IgTeCe*

Les travaux sur le système d'expression contenant *IgTeCe* ont été réalisés sur deux durées d'induction de 24 et 60 heures.

Induction de 24 heures

L'expression d'*IgTeCe* a été testée pour un temps d'induction de 24 heures. Les profils électrophorétiques par SDS-PAGE des extraits protéiques ont été comparés entre le témoin LacZ induit et *IgTeCe* induit (Figure 43). L'extraction a été réalisée avec un tampon 50 mM phosphate de sodium, pH 8. Au poids moléculaire attendu de 46 kDa, un nombre important de bandes sont

présentes. Aucune bande supplémentaire ou traduisant une surexpression n'apparaît pour l'extrait protéique de l'échantillon *IgTeCe*.

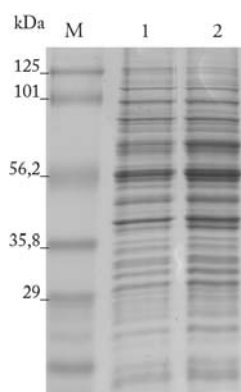


Figure 43: Electrophorèse sur SDS-PAGE 12% des extraits de protéines.

M : Marqueur de poids moléculaire, Piste 1: Extraits protéiques du témoin *LacZ*, Piste 2: Extraits protéiques *IgTeCe*.

Les extraits protéiques du témoin *LacZ* et d'*IgTeCe* ont ensuite été analysés par Western blot. La technique de SDS-PAGE n'a pas permis de visualiser la β -galactosidase issue du témoin *LacZ* (Figure 44 A). Cependant, l'analyse par Western blot a confirmé sa présence (Figure 44 B). Pour les échantillons provenant d'*IgTeCe*, aucun signal n'est observé ni en SDS-PAGE, ni en Western Blot (Figure 44 A et Figure 44 B).

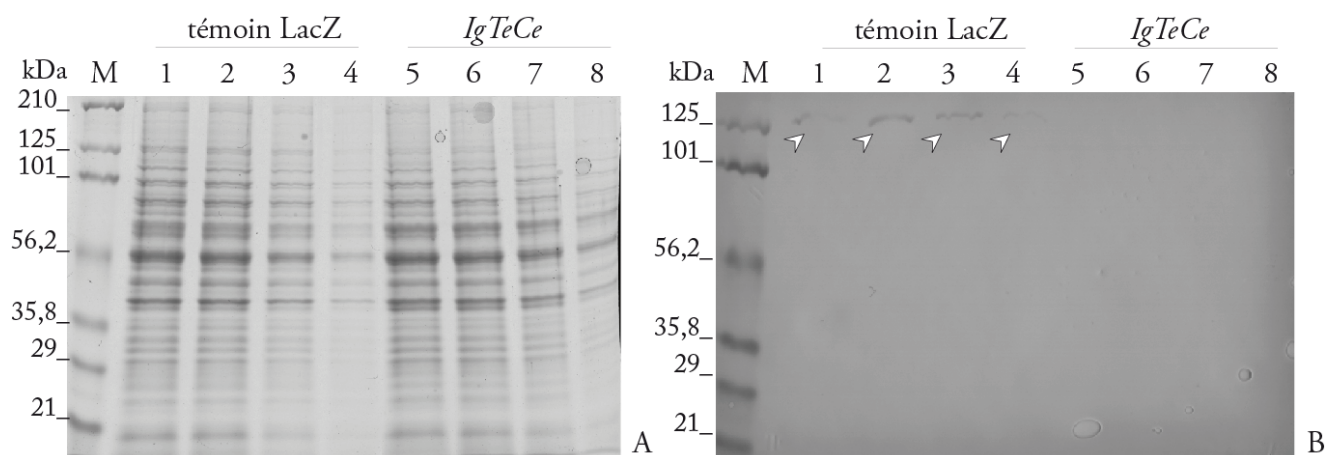


Figure 44: A) Electrophorèse sur SDS-PAGE 12 % des extraits protéiques déposés pour le Western blot. B) Western blot des extraits protéiques.

Témoin *LacZ* : M : Marqueur de poids moléculaire, Piste 1 : 20 μ g d'extrait protéique, Piste 2 : 15 μ g d'extrait protéique, Piste 3 : 10 μ g d'extrait protéique, Piste 4 : 5 μ g d'extrait protéique.

IgTeCe : Piste 5 : 20 μ g d'extrait protéique, Piste 6 : 15 μ g d'extrait protéique, Piste 7 : 10 μ g d'extrait protéique, Piste 8 : 5 μ g d'extrait protéique.

Différents tampons de lyse cellulaire ont été testés afin d'obtenir la protéine *IgTeCe*. Les profils d'électrophorèse se sont révélés identiques quel que soit le tampon d'extraction utilisé. C'est pourquoi le tampon phosphate de sodium 50 mM, pH 8 a été conservé.

L'analyse Western blot n'a pas permis de visualiser la protéine *IgTeCe*. Malgré la sensibilité de ce test, il est possible que la production de la protéine *IgTeCe* soit insuffisante pour permettre sa détection.

Des essais de chromatographie d'affinité ont été réalisés pour purifier la protéine *IgTeCe*. La séparation de la protéine d'intérêt par rapport aux autres protéines de la levure peut permettre de mieux distinguer sa production. Les fractions obtenues au cours de la chromatographie ont ensuite été analysées par SDS-PAGE.

Le résultat est présenté en figure 45 A et figure 45 B. Les pistes 1 et 2 correspondent respectivement aux extraits protéiques du témoin LacZ et d'*IgTeCe* avant dépôt sur la résine de chromatographie. La piste 3 correspond à la fraction non retenue. Son profil rassemble une grande partie des bandes observées sur l'extrait brut et particulièrement les protéines de poids moléculaire compris entre 29 et 100 kDa. La piste 4 correspond à la fraction de lavage de la résine. Peu de protéines ont été éluées de manière aspécifique. La piste 5 correspond à la première fraction éluee avec une concentration de 50 mM d'Imidazole. Aucune bande significative n'est visible. Les pistes 6 et 7 correspondent aux fractions éluées avec une concentration d'Imidazole de 100 mM. La piste 8 correspond à la fraction éluee à une concentration de 300 mM d'Imidazole. Pour ces trois pistes, une bande protéique de poids moléculaire proche de 56 kDa est détectée. Cette protéine n'étant pas détectée sur les autres pistes, il s'agit d'une protéine éluee par compétition.

La protéine éluee a un poids moléculaire trop élevé pour correspondre à la protéine *IgTeCe* sous sa forme native. Deux hypothèses sont envisageables. Une augmentation du poids moléculaire peut s'expliquer par une modification post-traductionnelle. Néanmoins, si l'étiquette poly-histidine est bien présente, elle aurait dû être détectée par Western blot. La protéine éluee peut également correspondre à une protéine de la levure, immobilisée sur la colonne de chromatographie de manière aspécifique puis éluee. Cette hypothèse nécessite de comparer les profils électrophorétiques entre les échantillons LacZ et *IgTeCe*.

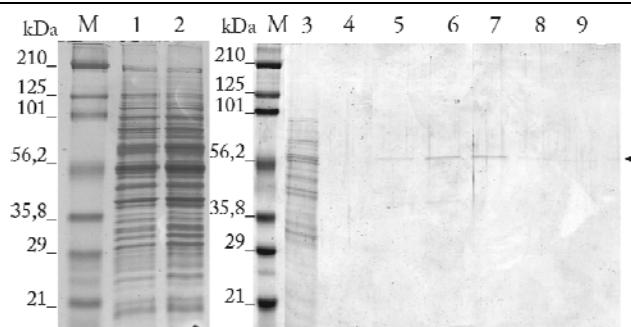


Figure 45: Electrophorèse sur SDS-PAGE 12 % des différentes fractions de chromatographie d'affinité pour *IgTeCe* après induction de 24 heures.

A) M : Marqueur de poids moléculaire, Piste 1 : Témoin LacZ, Piste 2 : Extrait protéique *IgTeCe*.

B) M : Marqueur de poids moléculaire, Piste 3 : Fraction non retenue *IgTeCe*, Piste 4 : Lavage de la colonne de chromatographie *IgTeCe*, Piste 5 : Elution à 50 mM d'Imidazole *IgTeCe*, Piste 6 : Elution à 100 mM d'Imidazole *IgTeCe*, Piste 7 : Elution à 100 mM d'Imidazole *IgTeCe*, Piste 8 : Elution à 300 mM d'Imidazole *IgTeCe*, Piste 9 : Elution à 500 mM d'Imidazole *IgTeCe*.

Afin d'optimiser l'obtention de la protéine élue, les volumes d'élution ont été réduits et la concentration d'Imidazole a été fixée à 150 mM. Un protocole de concentration et de dialyse a également été mis en place. L'objectif est d'augmenter la concentration de la protéine putative puis de comparer son profil d'électrophorèse à ceux de fractions du témoin LacZ. Les extraits obtenus sont comparés par SDS-PAGE et Western blot. La figure 46 A présente le SDS-PAGE. Les pistes 1 et 2 correspondent au témoin LacZ induit et la piste 3 correspond à *IgTeCe* induit. La figure 46 B présente le Western blot correspondant. Le Western blot permet de contrôler la présence d'un signal pour une protéine élue au cours d'une chromatographie d'affinité spécifique d'une étiquette poly-histidine. L'alignement des profils d'électrophorèse montre sur le profil du témoin LacZ une bande identique à celle observée pour *IgTeCe* lors de la chromatographie (indiqué par une flèche). Il s'agit donc d'une protéine de la levure retenue de manière aspécifique sur la résine de chromatographie. Le résultat du Western blot confirme ce résultat. Aucun signal n'est visualisé pour *IgTeCe* malgré la concentration de l'échantillon. Un signal est bien observé pour le témoin LacZ.

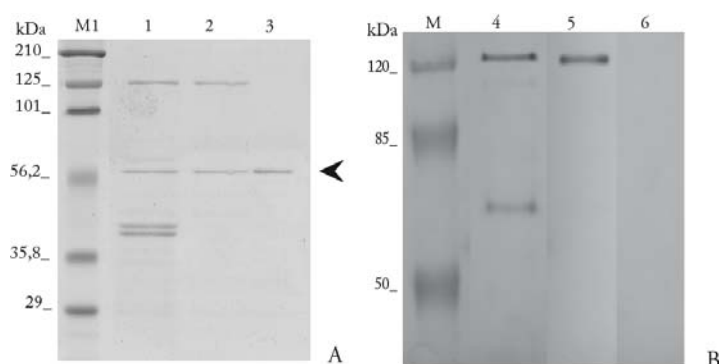


Figure 46: Comparaison entre le témoin LacZ et *IgTeCe*

A) SDS-PAGE 12 %. B) Western blot. M : Marqueur de poids moléculaire, Pistes 1, 4 : Témoin LacZ élué à 150 mM d'Imidazole, Pistes 2, 5 : Témoin LacZ élué à 500 mM d'Imidazole, Pistes 3, 6 : Fraction de chromatographie concentré et dialysée *pYE-IgTeCe* issue du regroupement de différentes éluions.

Induction de 60 heures

De nouveaux essais ont été réalisés après 60 heures d'induction. La chromatographie d'affinité a été réalisée et analysée par SDS-PAGE (Figure 47 A) et Western blot (Figure 47 B). La β -galactosidase est bien visible sur le profil électrophorétique du témoin LacZ à partir de l'extrait brut (flèche 1). Le fonctionnement de la chromatographie d'affinité est vérifié pour la β -galactosidase dont l'élution est réalisée à 50, 100 et 300 mM d'Imidazole (flèche 2, 3 et 4). La protéine *IgTeCe* reste indécélable sur gel.

L'analyse par Western blot a été réalisée pour les mêmes échantillons et a permis la détection de la β -galactosidase dans l'extrait brut (flèche 5) et dans chacune des fractions précédemment citées (flèche 6, 7 et 8). Aucun signal n'a été détecté pour *IgTeCe*.

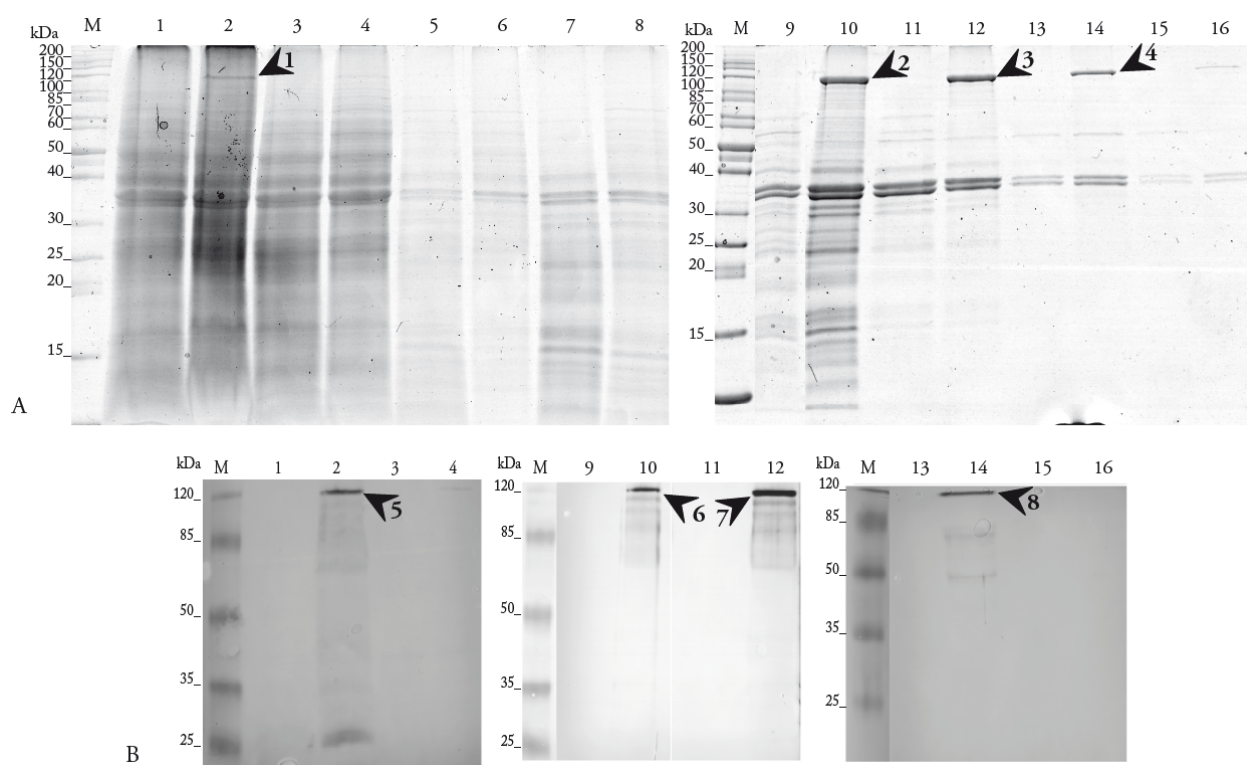


Figure 47: Analyse des fractions de chromatographie et des extraits protéiques obtenus pour le témoin LacZ et *IgTeCe* avec une induction de 60 heures. A) SDS-PAGE 12 % B) Western blot.

M : Marqueur de poids moléculaire, Piste 1 : Extrait protéique *IgTeCe*, Piste 2 : Extrait protéique témoin LacZ, Piste 3 : Fraction non retenue *IgTeCe*, Piste 4 : Fraction non retenue témoin LacZ, Piste 5-7 : Lavage de la colonne de chromatographie *IgTeCe*, Piste 6-8 : Lavage de la colonne de chromatographie témoin LacZ, Piste 9 : Elution à 50 mM d'Imidazole *IgTeCe*, Piste 10 : Elution à 50 mM témoin LacZ, Piste 11 : Elution à 100mM d'Imidazole *IgTeCe*, Piste 12 : Elution à 100 mM d'Imidazole témoin LacZ, Piste 13 : Elution à 300 mM d'Imidazole *IgTeCe*, Piste 14 : Elution à 300 mM d'Imidazole témoin LacZ, Piste 15 : Elution à 500 mM d'Imidazole *IgTeCe*, Piste 16 : Elution à 500 mM d'Imidazole témoin LacZ.

Les résultats des Western blot et de la chromatographie d'affinité n'ont pas permis de visualiser une bande protéique pouvant correspondre à la protéine *IgTeCe*. Une analyse par Dot blot a été réalisée

pour chaque étape d'extraction protéique. Le Dot blot s'est révélé négatif pour *IgTeCe*. La problématique observée pour *IgEst3* semble donc se reproduire pour l'expression d'*IgTeCe*. Les hypothèses pouvant expliquer ces résultats sont identiques : absence ou trop faible production de la protéine.

3.2.6. Recherche de transcrits *IgTeCe*

Les essais d'expression d'*IgTeCe* n'ont pas permis de détecter la protéine d'intérêt sur SDS-PAGE ou par Western blot, quel que soit le temps d'induction testé. Pour identifier si le problème est lié à l'étape de transcription (construction non fonctionnelle) ou à l'étape de traduction, une recherche de transcrits a été réalisée pour chaque temps d'induction (Figure 48). L'étape de retrotranscription a été menée avec des Oligo(dT) (Pistes 1 et 2) ou avec un couple d'amorces spécifiques d'*IgTeCe* (Pistes 3 et 4). Pour la culture de 60 heures, un contrôle positif de la réaction PCR (Piste 5) et un contrôle de non-contamination de l'extrait d'ARN total (Piste 6) ont été ajoutés.

Aucune bande spécifique n'est observée ni pour un temps d'induction de 24 heures (Figure 48 A), ni pour une induction de 60 heures (Figure 48 B). En l'absence de transcrits lors de l'induction, le problème serait lié à la construction.

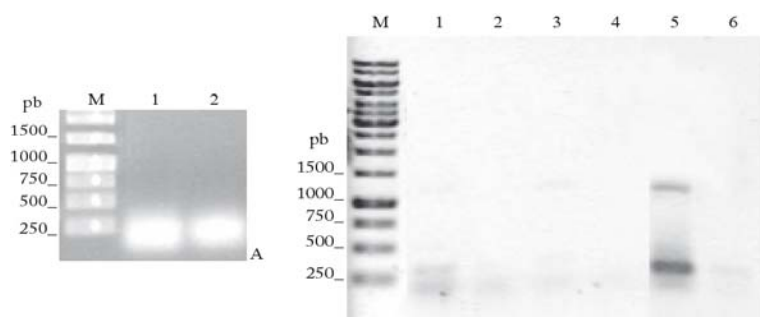


Figure 48: Electrophorèse sur gel d'agarose pour la recherche de transcrits *IgTeCe*.

La comparaison est réalisée entre les échantillons induits et les non-induits. A) Recherche réalisée pour une induction de 24 heures. B) Recherche réalisée pour une induction au bout de 60 heures. M : Marqueur de taille, Piste 1 : Non induit *IgTeCe*, Piste 2 : Induit *IgTeCe*, Piste 3 : Non induit *IgTeCe*, Piste 4 : Induit *IgTeCe*, Piste 5 : Contrôle positif de l'étape de PCR, Piste 6 : Contrôle de non contamination des ARN totaux induits.

3.2.7. Recherche d'activités lipolytiques pour *IgTeCe*

Pour le test d'activité lipolytique, les extraits protéiques ont été déposés sur un milieu gélosé contenant 10 % de tributyrine. La préparation enzymatique d'Amano, lipase F, a été utilisée comme contrôle positif. Le test a été réalisé pendant 120 heures à la température de 30°C. Afin d'évaluer jusqu'à quelle quantité de préparation enzymatique la dégradation du substrat est visible, la préparation enzymatique a été déposée jusqu'à une quantité théorique de 0,005 ng. Différents extraits d'*IgTeCe* ont été testés : l'extrait brut concentré, une fraction éluée à 50 mM, l'extrait brut non concentré et enfin un extrait brut issu du témoin LacZ et contenant la β -galactosidase.

Le résultat de l'ensemble des échantillons testés est négatif. L'activité d'hydrolyse de la lipase F a été observée au bout de 120 heures jusqu'à une quantité de préparation enzymatique égale à 5 ng (Figure 49).

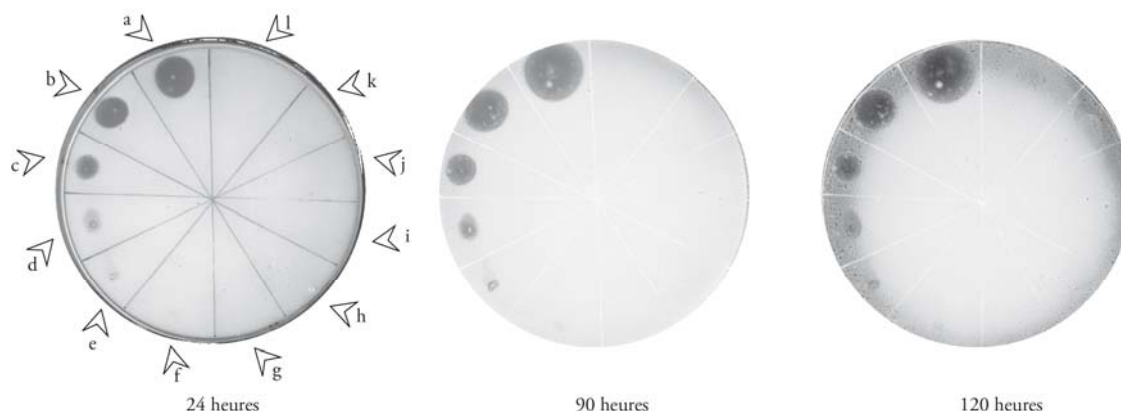


Figure 49: Test d'hydrolyse sur milieu gélosé avec 10 % de tributyrine.

a) 50 µg Lipase F, b) 5 µg Lipase F, c) 0,5 µg Lipase F, d) 50 ng Lipase F, e) 5 ng Lipase F, f) 0,5 ng Lipase F, g) 0,05 ng Lipase F, h) 0,005 ng Lipase F, i) Extrait brut concentré *IgTeCe*, j) Elution à 50 mM *IgTeCe*, k) Extrait brut *IgTeCe*, l) Extrait brut témoin *LacZ*.

3.2.8. Bilan des tests d'expression dans la levure

L'expression dans la levure du gène *IgTeCe* est comme dans le cas du gène *IgEst3* problématique. Malgré des conditions expérimentales permettant l'obtention d'une protéine témoin, la β -galactosidase, aucune production de la protéine *IgTeCe* n'a pu être mise en évidence.

3.3. Expression dans la bactérie *Escherichia coli*

L'expression dans la bactérie *Escherichia coli* vient compléter le travail réalisé dans la levure. Seule le clonage de la séquence codante *IgTeCe* a été obtenu. Ce travail de clonage a été effectué à partir de la construction utilisée lors de l'expression dans la levure de manière à isoler la séquence codante *IgTeCe*.

3.3.1. Clonage dans le plasmide *pJET1.2*

Dans un premier temps, à partir du plasmide *pYE-IgTeCe* présent dans *Saccharomyces cerevisiae*, la séquence codante *IgTeCe* a été isolée à l'aide du couple d'amorces 7519-ATG2 x 7519-STOP2E (suppression de l'alanine ajoutée lors du clonage dans la levure). La taille du fragment obtenu (1143 pb), appelé *IgTeCe-N*, correspond à celle attendue et a été validée par électrophorèse (Figure 50 A, piste 1).

Il reste cependant nécessaire de valider ce fragment par séquençage. Pour réaliser cette étape, le fragment *IgTeCe-N* a été inséré dans le plasmide *pJET1.2*. Ce plasmide va servir de vecteur de

stockage et faciliter le séquençage et la validation des inserts *IgTeCe* décrits dans la suite de ce paragraphe, en vue de leur sous-clonage dans les vecteurs d'expression.

Dans un second temps, les extrémités *d'IgTeCe-N* ont été modifiées afin de les rendre compatibles avec les vecteurs d'expression souhaités, *pET28a+* et *pET52b+*. Ces modifications par PCR ont été analysées par électrophorèse puis par séquençage et correspondent aux stratégies 1 et 2.

Stratégie 1 : clonage dans *pET28a+*

L'insertion d'*IgTeCe-N* dans *pET28a+* a nécessité l'ajout d'un site de restriction *NdeI* en 5' et d'un site de restriction *NotI* en 3'. Ce fragment sera appelé *IgTeCe-Nde-Not*. Le plasmide d'expression obtenu permettra la production d'une protéine *IgTeCe* disposant de deux étiquettes poly-histidine en C et N-terminal.

Stratégie 2 : clonage dans *pET52b+*

L'insertion d'*IgTeCe-N* dans *pET52b+* a nécessité l'ajout d'un site de restriction *NcoI* en 5' et d'un site de restriction *NotI* en 3'. Ce fragment sera appelé *IgTeCe-Nco-Not*. Le plasmide d'expression obtenu permettra la production d'une protéine *IgTeCe* disposant d'une étiquette poly-histidine en C-terminal.

Les fragments amplifiés ont été visualisés sur gel (Figure 50 B) avant insertion dans le vecteur *pJET1.2* et séquençage. La piste 1 correspond au fragment *IgTeCe-Nde-Not* et la piste 2 au fragment *IgTeCe-Nco-Not*.

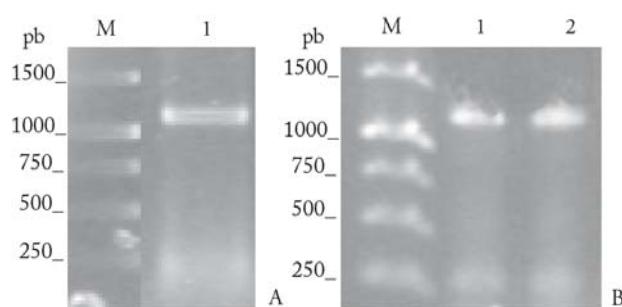


Figure 50: Electrophorèse sur gel d'agarose pour les amplifications de la séquence codante *IgTeCe*. A) M : Marqueur de taille, Piste 1 : Amplification avec le couple d'amorces 7519-ATG2 x 7519-STOP2E afin d'obtenir la séquence codante originale. B) M : Marqueur de taille, Piste 1 : Amplification avec le couple d'amorces 7519PTYFWNDE1 x 7519RVNOT1, Piste 2 : Amplification avec le couple d'amorces 7519FWNCO1 x 7519RVNOT1.

Résultat de séquençage pour *IgTeCe-N* (Annexe 9)

Les résultats de séquençage montrent que d'une part la séquence codante *IgTeCe-N* clonée est complète et d'autre part que celle-ci a été correctement intégrée dans le plasmide de clonage *pJET1.2*.

Résultat de séquençage pour IgTeCe-Nde-Not (Annexe 10)

Les résultats du séquençage montrent que l'ajout du site de restriction (*NdeI*) en 5' de la séquence codante n'a pas fonctionné. A contrario, le site de restriction (*NotI*) en 3' a bien été intégré. Néanmoins, la construction n'étant pas validée, elle ne sera pas utilisée pour le sous-clonage dans les vecteurs d'expression. L'ajout des extrémités à la séquence codante *IgTeCe* a été testé à plusieurs reprises sans succès.

Résultat de séquençage pour IgTeCe-Nco-Not

Aucun résultat de séquençage n'a été obtenu pour cette construction.

En conclusion, le clonage de la séquence codante d'*IgTeCe* est confirmé et la séquence codante native est disponible dans *pJET1.2*. Au vu des stratégies 1 et 2 infructueuses, l'utilisation des sites de restrictions disponibles sur le plasmide *pJET1.2* a été envisagé (Geszvain and Tebo, 2010). Le séquençage de différents clones a permis de sélectionner deux constructions contenant la séquence codante *IgTeCe-N*: *pJET-Ex1* et *pJET-Ex2* (Figure 51). *pJET-Ex1* est un plasmide contenant la séquence codante d'*IgTeCe* avec le site de restriction *NotI* en 5' et le site de restriction *XbaI* en 3'. *pJET-Ex2* est un plasmide contenant la séquence codante d'*IgTeCe* avec le site de restriction *NcoI* en 5' et le site de restriction *NotI* en 3'.

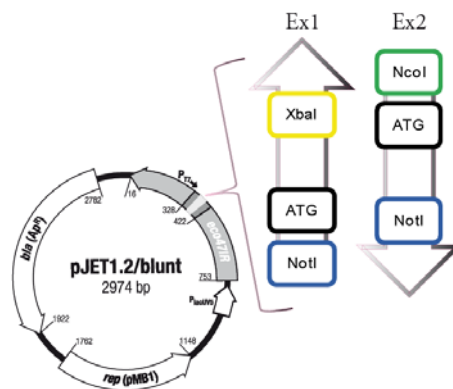


Figure 51: Illustration du produit de clonage d'*IgTeCe* obtenu et inséré dans pJET. Les séquençages complets des insertions sont présentés dans l'annexe 9.

Les constructions *pJET-Ex 1* et *pJET-Ex 2* vont donc servir pour le sous-clonage dans les vecteurs d'expression. La construction *pJET-Ex 2* permet notamment un sous clonage dans les vecteurs d'expression *pET28a+* et *pET52b+*. Dans ces conditions, les plasmides d'expression *pET28a+* et *pET52b+* doivent permettre la production d'une protéine *IgTeCe* disposant d'une étiquette poly-histidine en C –terminal.

3.3.2. Sous-clonage dans les vecteurs d'expression

Après sous clonage dans les vecteurs d'expression, les constructions ont été validées par séquençage (Annexes 11 et 12).

Chaque construction permet théoriquement l'obtention d'une protéine *IgTeCe* recombinante spécifique dont les caractéristiques sont présentées dans le tableau 60.

Nom de la construction	Propriétés de la protéine native	Propriétés de la protéine recombinante
<i>pET28-IgTeCe</i>	381 aa 42,31 kDa pI 9,8	404 aa 44,96 kDa pI 9,27
<i>pET52-IgTeCe</i>		417 aa 46,2 kDa pI 9,37

Tableau 60: Comparaison des modifications apportées à la protéine *IgTeCe* sous sa forme recombinée.

3.3.3. Expression hétérologue de la protéine *IgTeCe*

Deux souches bactériennes ont été utilisées :

- La souche Rosetta™ 2 avec les constructions *pET28-IgTeCe* et *pET52-IgTeCe*.
- La souche Star™ 2 uniquement avec la construction *pET28-IgTeCe*.

Les vecteurs d'expression disposent d'un système d'induction placé sous le contrôle de l'IPTG. Cet inducteur est un élément clef de la modulation de l'expression des protéines. C'est un analogue de l'allolactose non-métabolisable par la bactérie et dont la concentration reste constante dans le milieu de culture tout au long de l'induction.

Etude préliminaire de l'expression de la protéine IgTeCe par analyse des extraits bruts par SDS-PAGE et Western blot

Cette première approche a pour objectif de comparer différents profils électrophorétiques et de sélectionner l'association vecteur/souche/concentration d'IPTG la plus favorable. Deux concentrations en IPTG (0,4 et 1 mM) ont été testées. Ce test d'expression a été réalisé pour un temps d'induction de 6 heures (Yildir *et al.*, 1998). En complément, une chromatographie d'affinité a été effectuée pour *pET28-IgTeCe* et *pET52-IgTeCe*. L'analyse sur SDS-PAGE des extraits bruts et des fractions de chromatographie n'a pas permis de distinguer de différences notables entre les différentes conditions testées.

Une analyse complémentaire par Western blot a été réalisée. La première construction testée a été *pET28-IgTeCe* dans la cellule hôte Rosetta™ 2 (Figure 52). Le tampon de lyse utilisé est le tampon 100 mM Tris HCl, 500 mM KCl, 5 mM 2-mercaptoéthanol et 1 mM benzamidine, pH 8 (tampon de lyse 1).

Un signal a été obtenu pour le contrôle positif (piste 1), validant la méthode. Une bande majeure pour l'extrait *IgTeCe* induit à 1 mM (piste 2) est observée au niveau du marqueur de poids moléculaire de 50 kDa et une bande moins intense est observée entre 35 et 50 kDa. Cette bande protéique n'est observée que pour l'échantillon induit à 1 mM d'IPTG. Aucune bande n'a été détectée pour la culture non induite (piste 3). Aucun signal significatif n'a été détecté pour l'échantillon induit à 0,4 mM d'IPTG (résultat non présenté).

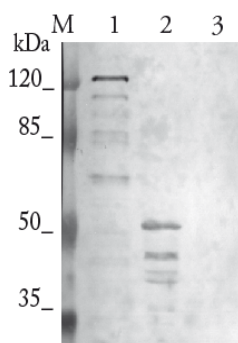


Figure 52: Western blot *pET28-IgTeCe* / Rosetta™ 2

Extraits protéiques obtenus après culture de 6 heures à 37 °C. M: Marqueur de poids moléculaire.

Piste 1: Contrôle positif, Piste 2 : Extrait brut *IgTeCe* induit à 1 mM, Piste 3: Extrait brut *IgTeCe* non induit.

Pour la construction *pET28-IgTeCe*, le culot de débris cellulaires après sonication a également fait l'objet d'une extraction avec le tampon de lyse 1 contenant 8 M d'urée (Figure 53). Pour cet extrait, un signal Western blot pour les échantillons induits à 0,4 mM (piste 3) et 1 mM d'IPTG (piste 2) a été observé. Ceci indique qu'à la concentration la plus faible en IPTG, la protéine détectée par Western blot a été précipitée dans le culot lors de l'expression.

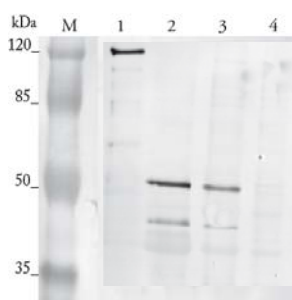


Figure 53: Western blot après extraction à 8M urée.

M: Marqueur de poids moléculaire, Piste 1: β -galactosidase, Piste 2 : Induit 1 mM, Piste 3: Induit 0,4 mM
Piste 4: Non induit.

D'autres essais de cultures ont été effectués à 37 °C avec un temps d'induction plus importants (8 heures) (Manderson *et al.*, 2006). Cet essai n'a pas été concluant pour *pET28-IgTeCe* par analyse Western blot.

Western blot appliqué aux autres constructions plasmidiques

Au vu du résultat du premier Western blot, l'expérience a été entreprise pour les systèmes d'expression *pET52-IgTeCe / Rosetta™ 2*, *pET28-IgTeCe / Rosetta™ 2* et *pET28-IgTeCe / Star™ 2* avec un temps de culture de 3 h à 37°C. Chaque analyse est composée d'un échantillon non induit et d'un triplicat d'échantillons induit avec 1 mM d'IPTG (Figure 54).

A. Evaluation de l'expression dans la souche Star™2.

Dans cette souche, l'expression basale n'est pas inhibée, il est donc possible d'observer un signal en absence d'IPTG. C'est effectivement ce qui a été observé pour le système *pET28-IgTeCe / Star™ 2* avec un signal à un poids moléculaire proche de 50 kDa (piste 1). Pour chacun des triplicats, une bande identique à la piste 1 est obtenue à un poids moléculaire proche de 50 kDa (pistes 2 à 4, figure 54 A). Ce poids moléculaire correspond à celui attendu pour la protéine *IgTeCe*. Cependant, une différence de niveau d'expression était attendue entre non induit et induit, le niveau d'expression pour les échantillons induits devant être plus important (largeur de la bande observée et intensité du signal). Ce n'est pas le cas. En conclusion, aucune surexpression n'a été détectée.

B. Evaluation de l'expression dans la souche Rosetta™ 2.

Dans cette souche, l'expression basale est inhibée, les profils doivent donc être différents entre les échantillons non induit et induit. Aucun signal n'est détecté aux pistes 1 et 5 correspondant aux échantillons non induits. L'analyse Western blot a révélé deux signaux pour le triplicat induit obtenu du système *pET52-IgTeCe / Rosetta™ 2* (pistes 2 à 4, figure 54 B). Le même résultat a été observé pour le système *pET28-IgTeCe / Rosetta™ 2* (piste 6 à 8, figure 54 B). Le signal majoritaire, détecté à un poids moléculaire d'environ 50 kDa, peut correspondre à la protéine *IgTeCe*. Le second signal, plus faible, détecté à environ 60 kDa pourrait correspondre à la protéine *IgTeCe* phosphorylée. La phosphorylation est, en effet, une des seules modifications post traductionnelles possibles chez la bactérie. Ce phénomène semble dépendre de la souche utilisée, car le signal observé est très faible au même poids moléculaire pour l'expression dans la souche Star™2 (Figure 54 A).

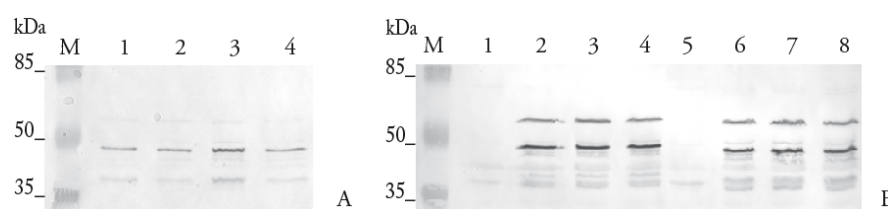


Figure 54: Validation de l'ensemble des constructions par Western blot

A) M : Marqueur de poids moléculaire, Piste 1 : Non induit *pET28-IgTeCe/Star™ 2*, Pistes 2 à 4 : Induit *pET28-IgTeCe/Star™ 2*. B) M : Marqueur de poids moléculaire, Piste 1 : Non induit *pET52-IgTeCe/Rosetta™ 2*, Pistes 2 à 4 : Induit *pET52-IgTeCe/Rosetta™ 2*, Piste 5: Non induit *pET28-IgTeCe/Rosetta™ 2*, Pistes 6 à 8 : Induit *pET28-IgTeCe/Rosetta™ 2*

En conclusion, pour chaque construction, un produit protéique est majoritairement détecté au poids moléculaire attendu après induction avec 1 mM IPTG. Pour l'expression dans la souche Rosetta™ 2, les extraits protéiques issus de *pET52-IgTeCe* et *pET28-IgTeCe* présentent des profils identiques par Western blot.

Effet des températures de culture sur l'expression de la protéine IgTeCe

Malgré l'observation d'un signal au poids moléculaire attendu, l'objectif de surexpression de la protéine *IgTeCe* n'a pas été atteint. Dans ce contexte, la littérature suggère de comparer les résultats en faisant varier les températures d'expression et en adaptant le temps d'induction à chaque température testée. Trois couples « température/durée d'induction » ont été testés : 3 heures pour 37 °C, 10 heures pour 30 °C et 16 heures pour 16 °C (Terpe, 2006; Tsuya *et al.*, 2006).

Lors d'une expression hétérologue, il est possible que la protéine forme des corps d'inclusion. Un moyen de diminuer ce phénomène est de travailler à des températures d'induction inférieures à 37 °C (Margreiter *et al.*, 2008). Afin d'étudier cet effet sur l'expression d'*IgTeCe*, trois températures ont été testées : 16, 30 et 37 °C. Deux tampons ont été utilisés pour l'extraction des protéines, un tampon PBS et un tampon PBS/8 M urée. Ce dernier devrait permettre l'extraction des protéines provenant de corps d'inclusion. L'analyse est réalisée par Western blot. La souche Rosetta™ 2 est utilisée de manière à avoir un contrôle non induit négatif.

Les résultats obtenus pour *pET52-IgTeCe* sont présentés en figure 55 (A et B). Les résultats obtenus pour *pET28-IgTeCe* sont présentés en figure 55 (C, D, E et F).

Des différences apparaissent en fonction de la construction choisie et de la température.

- Pour *pET52-IgTeCe*

Un signal est observé après expression à 16 °C (pistes 2 à 4, figure 55 A). Un faible signal est détecté pour l'extraction des corps d'inclusion (pistes 2 à 4, figure 55 B). Comparativement, le signal observé lors de l'expression à 30 °C (pistes 6 à 8, figure 55 A) n'excède pas celui observé à 37 °C (piste 10, figure 55 A). Les signaux détectés dans les corps d'inclusion (pistes 6 à 8, piste 10, figure 55 B) n'ont pas mis en évidence un effet « température » sur la formation de corps d'inclusion lors de l'expression de la protéine *IgTeCe*. Les conditions d'expression conservées pour ce système sont 16 °C et 37 °C.

- Pour *pET28-IgTeCe*

Les signaux observés après expression à 16 °C et 37 °C sont comparables (pistes 2 à 4, figure 55 C, pistes 2 à 4, figure 55 E). Lors de l'expression à 16 °C, une majeure partie du signal est détectée dans

les corps d'inclusion (pistes 2 à 4, figure 55 D). L'expression à 30 °C ne semble pas fonctionnelle car le signal Western blot est faible (pistes 6 à 8, figure 55 C, pistes 6 à 8, figure 55 D).

En conclusion, la température d'induction de 30 °C donne de moins bons résultats à l'exception de la construction *pET52-IgTeCe* (pistes 6 à 8, figure 55 A). Il faut également noter que la bande supplémentaire à environ 60 kDa est présente pour la majorité des conditions d'expression. Seule la construction *pET28-IgTeCe* ne présente pas ce signal lors de l'expression à 16 °C (pistes 2 à 4, figure 55 C).

Ces tests ont permis de retenir deux températures d'expression, 16 °C et 37 °C, et de sélectionner le système *pET52-IgTeCe/Rosetta™ 2* pour la recherche de transcrits et d'activité enzymatique.

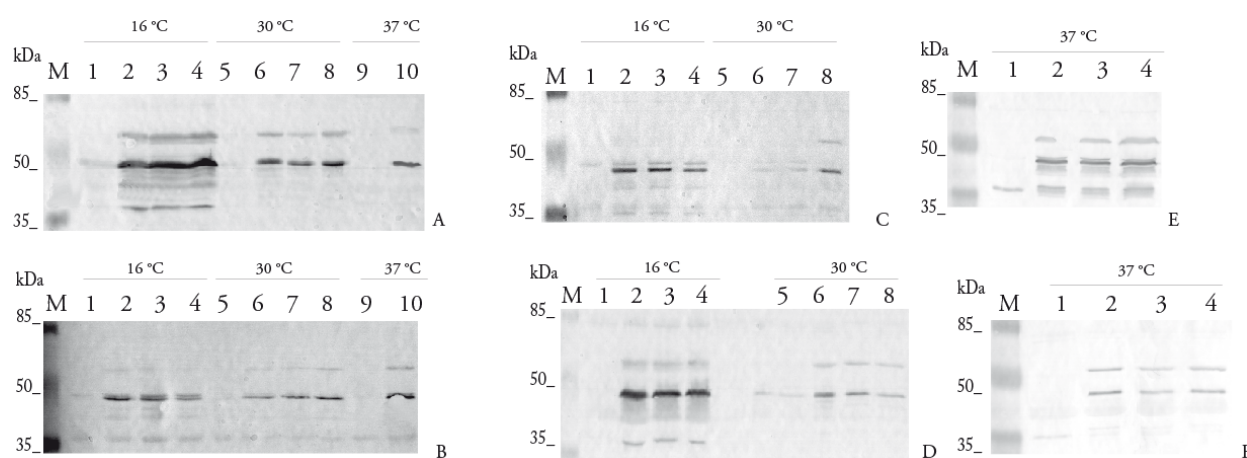


Figure 55: Western blot du test de différentes températures d'expression

A) *pET52-IgTeCe* / 16 – 30 – 37 °C / extraction classique, B) *pET52-IgTeCe* / 16 – 30 – 37 °C / extraction 8M urée, C) *pET28-IgTeCe* / 16 – 30 °C / extraction classique, D) *pET28-IgTeCe* / 16 – 30 °C / extraction 8M urée, E) *pET28-IgTeCe* / 37 °C / extraction classique, F) *pET28-IgTeCe* / 37 °C / extraction 8M urée.

3.3.4. Chromatographie d'affinité

L'étiquette poly-histidine, détectée par Western blot grâce à un anticorps spécifique, doit permettre la purification de la protéine *IgTeCe* par chromatographie d'affinité. La mise au point de cette purification a nécessité de choisir la résine et la composition du tampon utilisé à chaque étape (contact avec la résine, récupération de la fraction non retenue, lavage, les différentes fractions d'élution). Deux résines, la matrice « NiNTA » contenant l'ion divalent nickel et la matrice « TALON® » contenant l'ion divalent cobalt ont été testés. Dans un premier temps, le tampon 50 mM phosphate de sodium, 300 mM NaCl, pH8 a été choisi pour respecter les préconisations des fournisseurs de résine. Par la suite, parmi les différents tampons testés, seul le tampon PBS pH 8, servant à l'extraction des protéines, a été retenu dans les protocoles de purification. Chaque échantillon, après passage sur colonne de chromatographie d'affinité, a été analysé par SDS-PAGE et Western blot.

Les deux protéines provenant de *pET28-IgTeCe* et *pET52-IgTeCe* peuvent être purifiées par cette technique de chromatographie. La différence entre les deux protéines est le nombre de résidus histidine composant l'étiquette C-terminal (six pour *pET28-IgTeCe* et dix pour *pET52-IgTeCe*).

Après de nombreux essais, les résultats les plus exploitables sont obtenus à partir d'une culture induite de 3 heures à 37 °C avec 1 mM d'IPTG (Figure 56). Cependant, les différentes fractions éluées n'ont pas permis de purifier la protéine *IgTeCe* (pistes 4 à 12, figure 56 A et C). Pour apprécier l'effet de la chromatographie d'affinité, il est nécessaire d'analyser les différentes fractions par Western blot (Figure 56 B et D).

Pour *pET28-IgTeCe*, l'échantillon déposé est suffisamment concentré pour saturer la résine « TALON® ». En effet, la fraction non retenue présente un signal en Western blot comparable à celui de l'extrait brut (pistes 1 et 2, Figure 56 B). Un signal Western blot est obtenu dès la piste 4 correspondant à la première élution à 20 mM d'Imidazole. Cependant, ce signal n'est pas unique sur la membrane (piste 4). Des bandes sont détectées par l'anticorps au-dessus et au-dessous (flèches noires) de la bande de poids moléculaire attendu (flèche blanche). En piste 5 (2^e élution à 20 mM d'Imidazole), un signal de plus faible intensité est également visible au poids moléculaire attendu (flèche blanche), seule une très faible bande supplémentaire apparaît (flèche noire). L'analyse SDS-PAGE de ces mêmes pistes donne des résultats peu concluants. En effet, les profils SDS-PAGE obtenus sont trop complexes. Pour les autres pistes, le signal Western blot n'est plus détecté. Pour la piste 12 (Figure 56 A et B), le tampon de Laemmli (permettant de dénaturer la plupart des liaisons) a été utilisé directement sur la résine. La présence d'un signal révèle que le décrochage de la protéine d'intérêt n'a pas été complètement réalisé par compétition avec l'Imidazole.

Pour *pET52-IgTeCe*, l'échantillon déposé est également suffisamment concentré pour saturer la résine « NiNTA ». En effet, la fraction non retenue présente un signal en Western blot au poids moléculaire attendu (piste 2, Figure 56 D). En piste 4, le Western blot a révélé la présence de plusieurs protéines dont une, aspécifique, est proche de la bande d'intérêt (flèche noire du haut). En piste 5, la protéine d'intérêt est visualisée par un signal Western blot fort (flèche blanche), retrouvé plus faiblement en piste 6. Pour les pistes 4 et 5, des bandes supplémentaires sont détectées par Western blot (flèches noires). Par contre, l'analyse par SDS-PAGE des fractions éluées par chromatographie d'affinité ne permet pas de révéler une bande majeure correspondant à la protéine *IgTeCe* (Figure 56 C). L'étiquette de dix histidines semble néanmoins permettre une meilleure affinité et une élution plus

fine de la protéine d'intérêt. Un signal a également été observé en éluant avec le tampon de Laemmli sur la résine de chromatographie.

La chromatographie d'affinité n'a pas permis la purification de la protéine d'intérêt. Cependant, il faut noter que la longueur de l'étiquette poly-histidine semble être un facteur important. En effet, une meilleure rétention de la protéine sur la résine de chromatographie a été observée dans le cas de *pET52-IgTeCe*. A l'avenir, il pourrait être intéressant d'effectuer une étape de chromatographie supplémentaire avant la chromatographie d'affinité (chromatographie échangeuse d'ions par exemple). Ceci pourrait permettre d'éliminer une grande partie des protéines aspécifiques retrouvées dans nos analyses, et d'augmenter la pureté de la protéine d'intérêt.

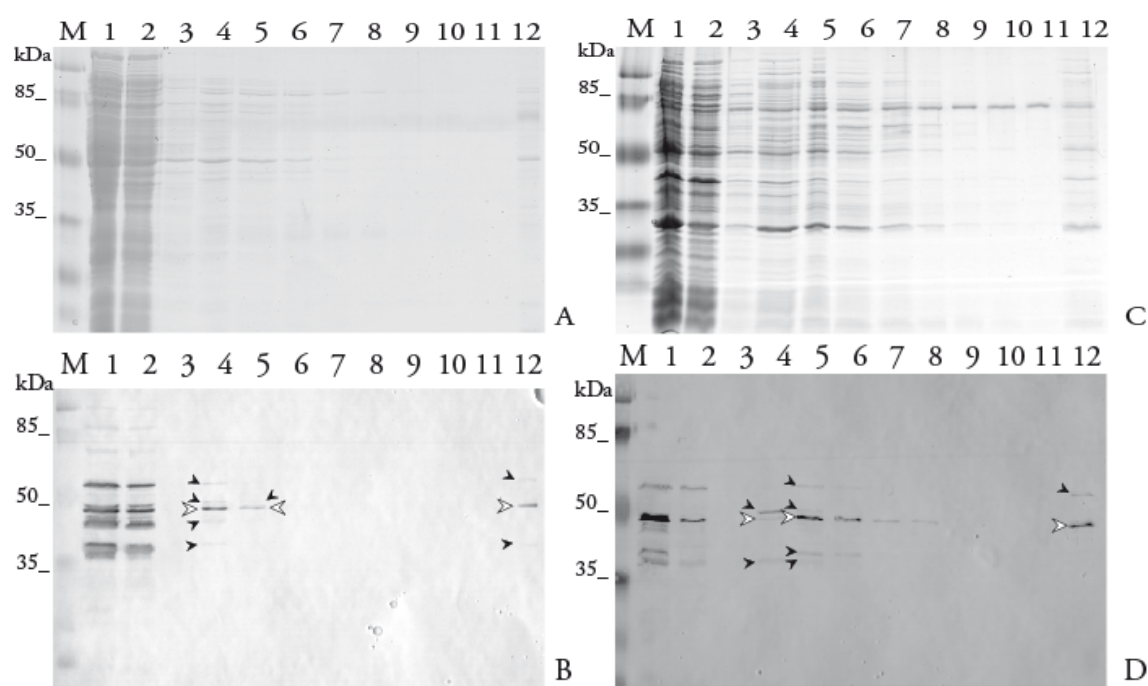


Figure 56: SDS-PAGE et Western blot des différentes fractions obtenues par chromatographie d'affinité.

A) SDS-PAGE 12 % des fractions de chromatographie sur colonne « TALON® » des extraits issus d'une culture avec *pET28-IgTeCe*

B) Western blot des fractions de chromatographie sur colonne « TALON® » des extraits issus d'une culture avec *pET28-IgTeCe*

M : Marqueur de poids moléculaire, Piste 1 : Extrait brut, Piste 2 : Fraction non retenue, Piste 3 : Fraction du lavage, Piste 4 : 1^{ère} élution 20 mM Imidazole, Piste 5 : 2^{nde} élution 20 mM Imidazole, Piste 6 : 3^{ème} élution 20 mM Imidazole, Piste 7 : 4^{ème} élution 20 mM Imidazole, Piste 8 : 1^{ère} élution 250 mM Imidazole, Piste 9 : 2^{nde} élution 250 mM Imidazole, Piste 10 : 3^{ème} élution 250 mM Imidazole, Piste 11 : 4^{ème} élution 250 mM Imidazole, Piste 12 : Elution tampon de Laemmli.

C) SDS-PAGE 12 % des fractions de chromatographie sur colonne « NiNTA » des extraits issus d'une culture avec *pET52-IgTeCe*

D) Western blot des fractions de chromatographie sur colonne « NiNTA » des extraits issus d'une culture avec *pET52-IgTeCe*

M : Marqueur de poids moléculaire, Piste 1 : Extrait brut, Piste 2 : Fraction non retenue, Piste 3 : Fraction du lavage, Piste 4 : 1^{ère} élution 50 mM Imidazole, Piste 5 : 2^{nde} élution 50 mM Imidazole, Piste 6 : 3^{ème} élution 50 mM Imidazole, Piste 7 : 4^{ème} élution 50 mM Imidazole, Piste 8 : 1^{ère} élution 250 mM Imidazole, Piste 9 : 2^{nde} élution 250 mM Imidazole, Piste 10 : 3^{ème} élution 250 mM Imidazole, Piste 11 : 4^{ème} élution 250 mM Imidazole, Piste 12 : Elution tampon de Laemmli.

3.4. Recherche de transcrits pour *IgTeCe* après expression dans la bactérie

Les ARN totaux sont extraits à la fois pour l'échantillon non induit et pour un duplicat d'échantillons induits. Une fois isolés et analysés, les ARN totaux sont soumis à une étape de retro-transcription en utilisant des amorces universelles. Les ADN complémentaires obtenus vont servir de matrice pour la réaction PCR. La séquence codante complète d'*IgTeCe* est amplifiée avec le couple d'amorces 7519-ATG2 x 7519-STOP2E.

Dans cette expérience, deux conditions de culture ont été testées pour la construction *pET52-IgTeCe*, l'une à 37 °C pendant 3 heures et l'autre à 16 °C pendant 16 heures. Elles correspondent à celles où le signal Western blot révèle la présence d'une protéine au poids moléculaire attendu.

Les résultats obtenus pour les échantillons des cultures induites à 37 °C pendant 3 heures sont présentés dans la figure 57. Pour l'échantillon non induit (piste 1), aucune bande pouvant correspondre à *IgTeCe* n'est observée. Pour les échantillons induits (pistes 2 et 3), une bande à la taille attendue pour *IgTeCe* est détectée. Les pistes 4, 5 et 6 correspondent à l'amplification des extraits d'ARN totaux. Aucun signal n'est visible, ces extractions n'ont donc pas été contaminées par de l'ADN plasmidique. La présence du transcrit est ainsi validée pour le système d'expression *pET52-IgTeCe*.

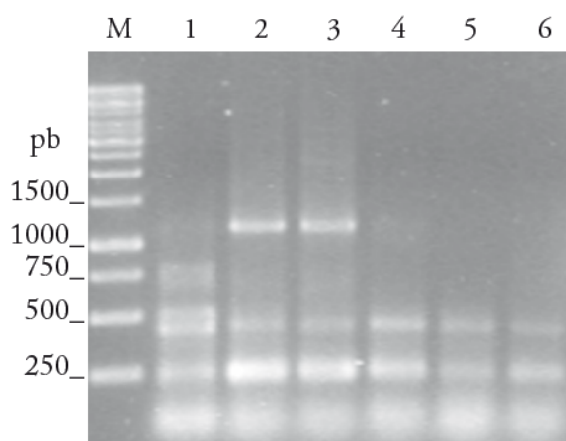


Figure 57: Résultat de l'étape de PCR pour la recherche de transcrits à 37°C.

M : Marqueur de taille, Piste 1 : PCR sur ADNc issus de l'échantillon non induit. Pistes 2 et 3 : PCR sur ADNc issus des échantillons induits, Piste 4 : PCR sur ARNtotaux issus de l'échantillon non induit, Pistes 5 et 6 : PCR sur ARNtotaux issus des échantillons induits.

Une recherche de transcrits a également été effectuée pour la culture induite à 16 °C pendant 16 heures (Figure 58). Un signal est observé pour les deux échantillons induits (pistes 2 et 3) et aucun signal n'est observé pour l'échantillon non induit (piste 1). Aucun signal n'est visible pour les pistes 4, 5 et 6, les ARN totaux n'ont pas été contaminés, la présence du transcrit *IgTeCe* est validée.

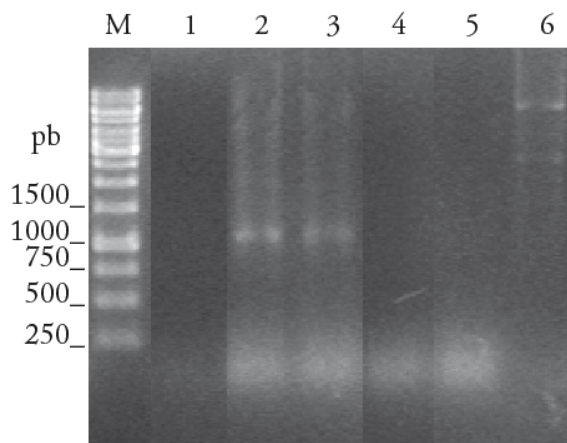


Figure 58: Résultat de l'étape de PCR lors de la recherche de transcrits à 16°C.

M : Marqueur de taille, Piste 1 : PCR sur ADNc issu de l'échantillon non induit. Pistes 2 et 3 : PCR sur ADNc issus des échantillons induits, Piste 4 : PCR sur ARN totaux issus de l'échantillon non induit, Pistes 5 et 6 : PCR sur ARN totaux issus des échantillons induits.

En conclusion, la recherche de transcrits s'est révélée positive avec des amorces spécifiques du gène *IgTeCe*. Le signal RT-PCR obtenu est donc le produit de l'amplification d'un ADNc correspondant à *IgTeCe*. L'utilisation d'amorces ciblant l'intégralité de la séquence a permis d'éviter une caractérisation d'un transcrit partiel. Ces tests viennent confirmer les résultats obtenus lors de la recherche de la protéine *IgTeCe* par Western blot.

4. Recherche d'activité pour *IgTeCe* après expression dans la bactérie

L'expression hétérologue a permis de détecter, pour la construction *pET52-IgTeCe*, un signal Western blot et un transcrit correspondant aux caractéristiques d'*IgTeCe*. A partir des clones positifs, l'extrait protéique obtenu a été utilisé comme catalyseur d'une éventuelle réaction lipolytique. L'activité est comparée par rapport à un témoin non induit. L'objectif est d'observer une activité enzymatique supérieure pour les échantillons induits par rapport au témoin, pouvant alors être mise en corrélation avec la présence de la protéine et du transcrit.

4.1. Hydrolyse de substrats chromophores

Les tests réalisés avec différents esters de para-nitrophényle se sont révélés négatifs pour l'acétate et l'octanoate.

4.2. Hydrolyse de substrats gélosés

Des tests sur des émulsions gélosées de différents substrats, avec comme contrôle positif la lipase F, ont été utilisés. Le dépôt d'échantillon a été effectué avec un volume constant contenant des quantités de protéines totales comparables. Le résultat est négatif pour la tributyrine (C4) mais également pour le Tween 20 : Monolaurate (C12) de polyoxyethylene-20-sorbitan. Concernant le Tween 80 : Monooléate (C18 :1) de polyoxyethylene-20-sorbitan, un changement d'aspect du milieu gélosé pour les échantillons induits a été observé.

Un autre substrat, la trioléine, a été testé sur milieu gélosé en microplaque. Chaque puits est composé d'une émulsion gélosée de trioléine à 2 % et complétée de rhodamine à 0,05 %. Une gamme de concentrations en protéines de chaque échantillon a été déposée (Induits en duplicat ; Non induit). Après une incubation de 24 heures à 37 °C, le test révèle une activité décelable à la fois pour l'induit et le non induit (Figure 59). Néanmoins, à 15 % d'extrait brut déposé, une activité lipolytique pour les échantillons induits est détectée. Pour 15 % d'extrait brut déposé issu d'un échantillon non induit, aucune activité n'est décelable. Ce résultat traduit une activité supérieure des échantillons induits sur la trioléine par rapport aux échantillons non induits. Ce résultat vient compléter le résultat précédent sur le Tween 80 (également C18 :1).

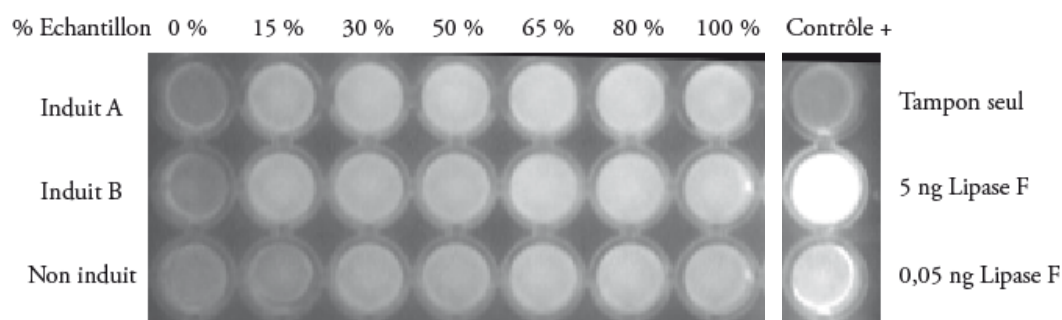


Figure 59: Activité lipolytique sur milieu gélosé contenant une émulsion de trioléine et l'indicateur rhodamine. Les extraits protéiques ont été obtenus après une culture induite à 37 °C pendant 3 heures.

Les substrats de longueurs de chaînes C2, C4 et C8 caractéristiques de l'activité carboxylestérase n'ont pas été hydrolysés par *IgTeCe*. La recherche d'une activité lipolytique a été élargie à des longueurs de chaînes plus importantes et seule une faible activité lipolytique a été supposée pour la trioléine C18 :1. La recherche d'activité étant négative pour la fonction carboxylestérase, l'analyse *in silico* a été reprise de manière à mieux préciser le type d'activité à attendre de la protéine *IgTeCe*.

4.3. Analyse de l'activité thioestérase

Pour rappel, les annotations originales (Banque TBestDB) et les premiers résultats de comparaison de séquences avec *IgTeCe* ont montré des homologies de séquences avec des lysophospholipases et des carboxylestérases. Par la suite, d'autres fonctions enzymatiques sont apparues telles que phospholipases et thioestérases. Malgré la conformation structurale du type hydrolases α/β suggérée par la structure secondaire et la présence d'un site consensus GX SXG, la protéine *IgTeCe* ne semble pas présenter d'activité carboxylestérase. Cette observation a conduit à s'intéresser à la fonction thioestérase.

Récemment, de nombreuses études se sont appuyées sur les enzymes du type thioestérase pour mettre en évidence l'enrichissement en lipides de microorganismes. Ces études ont montré un moyen de détecter l'activité d'une thioestérase produite de manière hétérologue en étudiant les variations de contenu en acide gras des lipides totaux de l'hôte. La littérature a notamment décrit, chez l'hôte *Escherichia coli*, une activité thioestérase pour une enzyme issue de la microalgue *Phaeodactylum tricornutum* (Gong *et al.*, 2011). C'est en se basant sur cette étude puis sur l'étude en 2012 d'une thioestérase d'*Acinetobacter baylyi* (Zheng *et al.*, 2012) et enfin sur l'étude de thioestérases de plantes (Zhang *et al.*, 2011), que l'analyse de l'effet de l'expression d'*IgTeCe* a été entreprise.

Une recherche de modification dans le profil d'acides gras des lipides totaux a été effectuée pour la construction *pET52-IgTeCe* présente dans la souche Rosetta™ 2. L'étude de l'activité thioestérase a

été réalisée à 16 et 37 °C. Dans chaque cas, un contrôle non induit n'exprimant pas la protéine *IgTeCe* a été utilisé comme témoin négatif.

Profil en acides gras pour les cultures à 16 °C

Après extraction des lipides totaux, l'analyse par chromatographie en phase gazeuse a permis d'établir le contenu en acides gras. Les profils sont comparés entre les échantillons induits et non induits.

L'acide gras majoritaire au sein des échantillons est l'acide oléique C18 :1, sa proportion ne varie pas entre les deux échantillons. Il représente 40 % des acides gras des lipides extraits. La comparaison de profils entre les échantillons induits et non induits révèle une différence majeure pour l'acide palmitoléique C16 :1 dont la proportion est quasiment doublée lors de l'induction (Figure 60). Le changement de profil concerne de manière plus limitée les autres acides. Deux pics n'ont pas été identifiés sur le profil d'acides gras. Ces acides gras non indentifiés ont notamment diminué dans l'échantillon induit. Ce résultat est à mettre en corrélation avec la présence d'un signal Western blot correspondant à la protéine *IgTeCe* et la caractérisation d'un transcrit spécifique de la séquence complète d'*IgTeCe*.

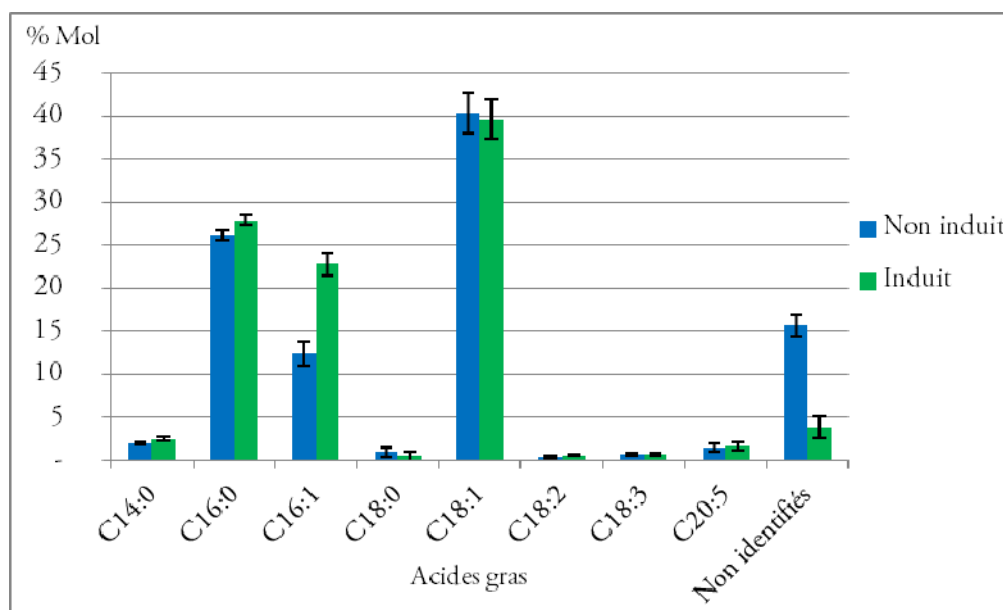


Figure 60: Composition du contenu en acides gras pour la culture à 16 °C.

Profil en acides gras pour les cultures à 37 °C

Les profils de contenu en acides gras des échantillons induits et non induits ont été comparés. L'acide gras majoritaire est l'acide palmitique C16 :0, il représente presque 45 % des acides gras des lipides extraits (Figure 61). La comparaison de profils entre les échantillons induits et non induits révèle un changement pour l'acide palmitoléique C16 :1 dont la proportion passe d'environ 2 % à 10 % en

présence d'*IgTeCe*. Une seconde modification concerne l'acide oléique (C18 :1) dont la proportion augmente de 15 à 20 %. Le changement de profil concerne de manière plus limitée les autres acides gras. Un certain nombre de pics sont également non identifiés sur le profil d'acides gras. Dans l'échantillon induit, ces pics diminuent.

Cette activité enzymatique complète la détection de la protéine et du transcrit spécifique d'*IgTeCe*.

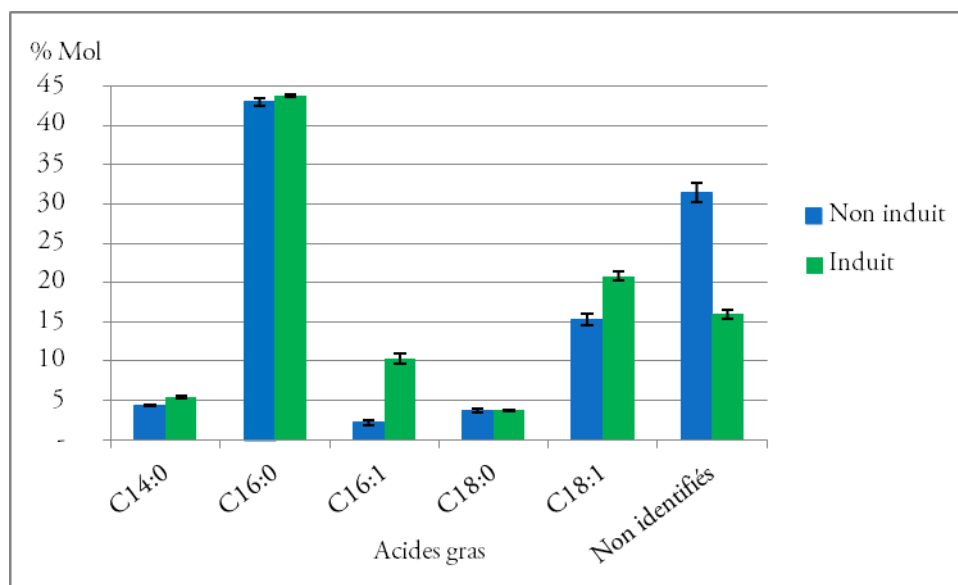


Figure 61: Composition du contenu en acides gras pour la culture à 37 °C.

Comparaison et bilan des résultats de CPG

Chacune des températures d'expression a abouti à l'observation d'un changement de profil en acides gras des lipides totaux de la cellule hôte.

La proportion de l'acide palmitoléique (C16 :1) augmente de $12,41 \pm 1,35$ % à $22,83 \pm 0,19$ % à la température de 16 °C. A 37 °C, la proportion est de $2,16 \pm 0,31$ % en absence de la protéine *IgTeCe* et de $10,34 \pm 0,65$ % en présence de la protéine *IgTeCe*. L'activité d'*IgTeCe* pour l'acide palmitoléique augmente d'un facteur 2 à 16 °C et d'un facteur 5 à 37 °C.

La proportion de l'acide oléique (C18 :1) n'évolue pas à 16 °C en présence ou absence de la protéine *IgTeCe*. En augmentant la température à 37 °C, la proportion de C18 :1 sans *IgTeCe* est de $15,35 \pm 0,75$ % et passe à $20,83 \pm 0,60$ % avec *IgTeCe*. La protéine *IgTeCe* n'impacte pas la proportion de C18 :1 à 16 °C. Une augmentation de la proportion de C18 :1 est par contre détectée à 37 °C.

La proportion des acides gras non identifiés à 16 °C est de $15,75 \pm 1,24$ % en l'absence d'*IgTeCe* et diminue à $3,84 \pm 0,06$ % en présence d'*IgTeCe*. A 37 °C la proportion des acides gras non identifiés est de $31,44 \pm 1,18$ %, elle diminue à $15,98 \pm 0,59$ % en présence de la protéine *IgTeCe*.

En conclusion, la protéine *IgTeCe* augmente la proportion de C16 :1 aux deux températures. Cette activité est complétée par l'augmentation de la proportion de C18 :1 lorsque la température passe à 37 °C. Il existe donc une corrélation température / spécificité de l'activité pour la protéine *IgTeCe*.

5. Discussion générale

Préalablement à ces travaux, le laboratoire avait mis en évidence une activité lipolytique endogène chez *Isochrysis galbana* (Poisson L, Thèse, 1999). Des activités de type phospholipase, glycolipase puis galactolipase avaient été supposées (Devos M, Thèse 2005). Une précédente étude avait également montré l'existence de dix gènes pouvant coder des enzymes lipolytiques dans cette microalgue. L'isolement et la caractérisation de la plupart de ces gènes avaient alors été initiés. Trois gènes avaient ainsi été étudiés, permettant l'identification d'une lipase *IgLip* et de deux carboxylestérases *IgEst1* et *IgEst2* (Godet *et al.*, 2010; Godet *et al.*, 2012). Dans ce contexte, l'objectif général des travaux de cette thèse est l'obtention, par expression hétérologue, de deux enzymes lipolytiques putatives *IgEst3* et *IgTeCe* codées par deux gènes de la microalgue *Isochrysis galbana*.

IgEst3 et *IgTeCe* correspondent respectivement aux marqueurs de séquences exprimées (cluster) ISL00006895 et ISL00007519 présents dans la base de données de séquences incomplètes, TBestDB (Keeling, 2006). A la constitution de cette banque pour *Isochrysis galbana*, une annotation automatique, permettant d'identifier d'un point de vue fonctionnel ces deux clusters, avait été réalisée. Ces annotations, obtenues par alignement des clusters, avaient permis de prédire que la protéine *IgEst3* appartenait à la classe d'ester hydrolase et pouvait avoir les fonctions de phospholipase ou de carboxylestérase et que la protéine *IgTeCe* pouvait avoir les fonctions de carboxylestérase et de lysophospholipase. Les travaux menés au laboratoire avaient permis d'isoler les séquences complètes incluant ces deux clusters. Ces séquences codantes avaient ensuite été clonées dans la levure *Saccharomyces cerevisiae*.

Au cours de cette étude, des analyses par comparaison de séquences ont permis de montrer de nombreuses homologies entre des enzymes lipolytiques et les séquences protéiques codées par ces deux gènes. En complément au système d'expression levure, la séquence codante d'*IgTeCe* a été

clonée dans la bactérie *Escherichia coli*. Différents tests d'expression ont abouti à la mise en évidence d'une activité lipolytique pour la protéine *IgTeCe*.

5.1. Evolution de prédictions des fonctions d'*IgEst3* et d'*IgTeCe*

Les travaux menés au laboratoire au cours de cette thèse ont débuté par l'évaluation des prédictions de fonction. Les alignements obtenus ont permis de mettre à jour les données et de confirmer ou redéfinir les fonctions enzymatiques prédites pour chaque protéine.

Pour la séquence protéique *IgEst3*, les algorithmes BLASTP ont majoritairement prédit les fonctions carboxylestérase et thioestérase. Pour ces deux fonctions, les pourcentages d'identité sont compris entre 23 et 35 % et les pourcentages de similarités sont compris entre 32 et 51 %. L'algorithme PipeAlign a classé la séquence *IgEst3* dans une nouvelle famille d'enzymes incluant également des thioestérases.

Pour la séquence *IgTeCe*, les algorithmes ont prédit une appartenance à la famille des hydrolases α/β montrant des homologies significatives avec les phospholipases, carboxylestérases, lysophospholipases et thioestérases. L'algorithme PipeAlign a classé la séquence *IgTeCe* dans une famille d'enzymes de fonction inconnue. PipeAlign n'a inclus aucune des fonctions proposées par BLASTP de manière significative dans la famille d'enzyme d'*IgTeCe*.

Les séquences homologues ayant permis l'annotation des clusters ont été comparées aux séquences complètes obtenues au laboratoire par l'algorithme BLASTP. Les scores et les pourcentages d'identités et de similarités ont été ainsi recalculés. Les données obtenues, à la fois pour la protéine *IgEst3* et la protéine *IgTeCe*, ne montrent pas de changements importants des pourcentages d'identité ou de similarité. En revanche, ces données montrent une évolution de l'annotation *in silico* d'*IgEst3*. La fonction carboxylestérase est conservée tandis que la fonction phospholipase disparaît au profit de la fonction thioestérase. Pour *IgTeCe*, ces données montrent un élargissement des fonctions possibles par rapport aux annotations initiales, ne facilitant pas sa caractérisation. En conclusion, il semble que l'enrichissement croissant des banques de données permette d'élargir le champ des fonctions enzymatiques possibles mais ne permette pas d'obtenir de séquences significativement plus homologues.

La démarche expérimentale d'annotation de séquences que nous avons utilisée pour définir les fonctions putatives d'*IgEst3* et d'*IgTeCe* est depuis quelques années en plein essor. En effet, le

développement rapide de nombreux programmes de séquençage a généré un grand nombre de séquences nucléotidiques et protéiques à caractériser. Dans notre étude, nous avons choisi de travailler avec l'algorithme BLASTP et le logiciel PipeAlign.

A partir des résultats obtenus par BLASTP, il est possible de définir une fonction enzymatique par homologie de séquence. L'alignement est évalué par un score et des pourcentages d'identité et de similarité (Altschul *et al.*, 1990). PipeAlign, d'une approche différente, permet d'identifier et de classer une séquence donnée dans une famille d'enzymes (Plewniak *et al.*, 2003).

Dans notre démarche, il faut noter les difficultés rencontrées par l'utilisation de l'algorithme de base BLASTP. En effet, en se basant sur les meilleurs scores obtenus, les résultats ont présenté des séquences peu informatives mais fortement homologues. Ce phénomène est dû à la présence de séquences non annotées ou non caractérisées dans les banques de données et nous a obligé à ignorer manuellement ces séquences non informatives dans notre démarche de prédiction de fonction. De la même manière, l'utilisation de l'algorithme dérivé PHI-BLAST n'a pas permis de restreindre le nombre de fonctions pour *IgEst3* ou *IgTeCe*. Dans cette stratégie d'alignement de séquence, seul l'algorithme DELTA-BLAST a permis de diminuer le nombre de séquences homologues peu ou pas informatives. En effet, il a généré une proposition de séquences homologues bénéficiant de scores et pourcentages d'identité et de similarité significatifs. Cette évolution de l'algorithme de base BLASTP génère tout d'abord une matrice de calcul par rapport à un domaine fonctionnel détecté sur la séquence candidate. Cette matrice sert ensuite de base aux calculs d'alignements de séquences (Boratyn *et al.*, 2012). Malgré l'obtention directe de séquences informatives, cet algorithme n'a pas permis de diminuer le nombre de fonctions enzymatiques possibles.

Ces résultats viennent confirmer ceux observés dans une revue qui s'est intéressée à l'évaluation des algorithmes d'analyse et de classement de séquences (Radivojac *et al.*, 2013). Plusieurs aspects sont décrits, notamment la diminution de l'efficacité de l'algorithme BLAST dans le contexte actuel où de nombreuses séquences sont disponibles mais non annotées dans les banques de données. L'évaluation de chaque algorithme est basée sur la capacité de celui-ci à prédire une fonction par la suite réellement observée d'un point de vue expérimental. Par cette méthode, il a été, par exemple, constaté que l'utilisation de l'algorithme PHI-BLAST ne permettait pas d'approfondir les résultats obtenus avec l'algorithme BLAST seul. C'est ce que nous avons pu constater avec l'analyse d'*IgEst3*. La méthodologie la plus propice à la prédiction d'une fonction enzymatique par BLASTP nécessite l'utilisation de plusieurs banques de données et privilégie notamment certaines d'entre elles où

l'annotation des séquences inventoriées a été réalisée manuellement comme par exemple la SwissProt. Dans le cas d'*IgTeCe*, cette banque de données a effectivement prédit une fonction thioestérase, correspondant à celle vérifiée expérimentalement. Pour *IgEst3*, l'absence de résultats expérimentaux probants n'a pas permis de confronter la prédiction obtenue avec la banque SwissProt.

Lors de la recherche par BLASTP, une détection de domaines protéiques vient compléter la recherche d'homologies. Pour *IgTeCe* comme pour *IgEst3*, cette prédiction a révélé un domaine de type α/β hydrolase en C-terminal. Ce domaine couvre les 2/3 de la séquence globale des protéines. Les parties N-terminal d'*IgEst3* et d'*IgTeCe* restent non identifiées dans les banques de données. Au cours de ce travail, seule la recherche de structures secondaires à l'aide de l'algorithme YASPIN, a permis d'obtenir des données et mis en évidence une suite d'hélices α , en N-terminal pour la séquence protéique d'*IgTeCe*. Toutefois, les parties N-terminal contiennent en général les sites de modifications post-traductionnelles ou des signaux peptidiques permettant l'orientation de la protéine vers son compartiment fonctionnel (Rusch and Kendall, 1995).

Parmi les séquences homologues à *IgTeCe*, la séquence de la microalgue *Emiliana huxleyi* présente également un domaine N-terminal inconnu et non homologue aux autres séquences alignées. L'absence d'identification de domaines connus et de signaux peptidiques détectés par le logiciel SignalP permet d'imaginer qu'il s'agit d'un domaine spécifique aux microalgues dont le rôle n'a pas encore été caractérisé.

De manière générale, les méthodes de prédictions de fonction par homologie de séquences aboutissent rarement à une seule sous classification EC x.x.x.x (Nagao *et al.*, 2014). Dans le cas des hydrolases α/β , la fonction catalytique est liée à la conservation de la structure tridimensionnelle de la protéine et non à la composition en acides aminés de la séquence protéique (Holmquist, 2000; Wong and Schotz, 2002). Par conséquent, la recherche d'une fonction enzymatique doit prendre en compte les homologies de séquence, les domaines fonctionnels et la prédiction de structures.

La majorité des informations que nous avons obtenues pour *IgEst3* et *IgTeCe* tendent à opposer les fonctions thioestérase et carboxylestérase. Cependant, les algorithmes de prédictions prévoient une conformation structurale du type hydrolases α/β avec l'enchaînement des structures en feuillet et en hélice. Si cette conformation peut correspondre à la fonction de carboxylestérase, elle est en revanche peu adaptée à une partie des thioestérases, notamment celles spécifiques des substrats de type Acyl-CoA qui adoptent une conformation du type « Hotdog » (Dillon and Bateman, 2004; Mayer and Shanklin, 2005, 2007; Cantu *et al.*, 2010). Néanmoins, ces oppositions et ces classifications

pour les thioestérases sont à nuancer. L'étude de Brocker *et al.* (2010) a montré que la famille des thioestérases humaines spécifiques de l'Acyl-CoA possèdent les deux conformations pour une même activité. De plus, les thioestérases, avec une conformation d'hydrolase α/β , présentent des séquences et des structures qui semblent conservées entre elles. Ils ont également montré la présence d'un domaine « lipase/estérase » en C-terminal, ce domaine est ou non précédé d'un domaine non structuré en N-terminal. Toutes ces informations permettent de conclure que la fonction thioestérase peut être compatible avec une conformation de type hydrolase α/β .

A ce stade, la prédiction des fonctions d'*IgEst3* et d'*IgTeCe* a permis de conforter l'appartenance à la super famille des hydrolases α/β . Néanmoins, l'association séquence/fonction obtenue par la recherche d'homologie de séquence demeure un outil fragile et ne peut rester le seul outil de qualification de la fonction d'une enzyme (à l'échelle de la sous-classe enzymatique). Il convient de lui préférer la recherche de domaines (Tian and Skolnick, 2003; Addou *et al.*, 2009).

Même si la recherche d'homologie ne nous a pas permis d'établir une fonction précise pour les protéines *IgEst3* et *IgTeCe*, elle a tout de même permis d'identifier les acides aminés (Ser, Asp et His) potentiellement impliqués dans leurs triades catalytiques. Cette détermination a par ailleurs été réalisée en utilisant des structures obtenues de carboxylestérases et de thioestérases présentant une conformation d'hydrolase α/β .

5.2. Expression hétérologue d'*IgEst3* et d'*IgTeCe*

5.2.1. Résultats de l'expression dans la levure *Saccharomyces cerevisiae*

Dans la littérature, l'hôte d'expression *S. cerevisiae* a été notamment utilisé pour l'étude de désaturases et d'élongases de microalgues ainsi que pour l'expression d'enzymes lipolytiques de diverses origines (Lopez-Lopez *et al.*, 2010; Li *et al.*, 2011; Shi *et al.*, 2012). L'expression d'*IgEst3* et d'*IgTeCe* dans la levure s'est donc appuyée sur certaines conditions décrites dans ces études. Plusieurs « couples » température/temps d'induction étaient envisageables. Par exemple, l'étude d'une carboxylestérase de *Thermus thermophilus* a été menée en exprimant la protéine durant 72 h à 30 °C, permettant ainsi la purification de la protéine et la caractérisation de son activité (Lopez-Lopez *et al.*, 2010). De la même manière, l'expression d'un gène de la microalgue *Phaeodactylum tricornutum* codant une acyl-transférase a été réalisée à 30 °C pendant 48 heures (Guiheneuf *et al.*, 2011). Notre choix s'est porté sur une température d'expression de 30 °C et des temps d'induction de 24 et 60 heures. Les conditions testées sont donc cohérentes avec celles retrouvées dans la littérature.

Cependant, ces conditions n'ont pas permis l'obtention d'une protéine détectable sur SDS-PAGE ou sur Western blot. Une recherche de transcrits a également été menée mais elle s'est révélée négative. L'absence totale de signe d'expression nous amène à formuler différentes hypothèses.

La première est la possible toxicité de la protéine produite. Cependant, il peut paraître étonnant que le résultat soit identique à la fois pour *IgEst3* et pour *IgTeCe*.

La seconde est l'incompatibilité des séquences codantes issues de microalgues pour leur expression dans la levure. Cependant, d'autres gènes issus de microalgues ont déjà été exprimés, notamment au laboratoire, permettant la détection de transcrits pour *IgEst1* et *IgEst2* (Godet S, Thèse 2008).

Enfin, il est également possible que les modifications apportées à la séquence native aient posé problème. L'ajout de l'acide aminé alanine en position 2 est censé assurer une augmentation de l'expression de la protéine (Tats *et al.*, 2006). Cette modification de la partie N-terminal, a pu être une source d'instabilité au moment de l'expression dans la levure et pourrait expliquer l'absence d'expression autant pour *IgEst3* et que pour *IgTeCe*. De plus, le manque d'informations concernant la fonction de la partie N-terminal pourrait également être à l'origine du défaut de production de la protéine.

Concernant le système d'expression, il a été testé sur le gène LacZ qui sert de témoin. La β -galactosidase a été produite et détectée sur SDS-PAGE. L'obtention de cette protéine a donc écarté l'hypothèse d'un problème de condition / milieu de culture / défaillance de la cellule hôte. En effet, la construction pour ce témoin est basée sur le vecteur d'expression pYES2.1 et la souche d'expression INVSc1 est identique à celle utilisée pour *IgEst3* et *IgTeCe*.

L'utilisation du témoin LacZ en parallèle des tests d'expression d'*IgEst3* et d'*IgTeCe* a cependant restreint les conditions expérimentales aux conditions où la β -galactosidase était obtenue. Au vu des résultats, il est possible que cette condition d'expression ne corresponde pas à la condition d'obtention des protéines *IgEst3* ou *IgTeCe*. En revanche, ce témoin a permis d'évaluer la capacité de surexpression dans les conditions utilisées lors de notre étude. L'expression de la β -galactosidase a été observée sur SDS-PAGE uniquement après 60 h d'induction par une bande caractéristique au poids moléculaire de 121 kDa. L'intensité de la bande ne traduit pas une « surexpression » notable. Le signal Western blot correspondant a été détecté dès 24 h d'induction. Ces résultats indiquent que d'autres conditions d'expression sont à tester, comme par exemple la température d'expression. Une élongase et une désaturase d'*Isochrysis galbana* ont été exprimées à une température de 22 °C (Qi *et*

al., 2002; Shi *et al.*, 2012). Cette méthode a été également utilisée pour une désaturase de la microalgue *Pavlova lutheri* (Tonon *et al.*, 2003).

5.2.2. Résultats de l'expression d'*IgTeCe* dans la bactérie *Escherichia coli*

Le clonage et l'expression d'*IgTeCe* dans la bactérie ont été réalisés de manière classique. La littérature est importante dans le cas de l'expression de lipases, de carboxylestérases ou de phospholipases (Hodgson *et al.*, 1990; Mnisi *et al.*, 2005). Dans un premier temps, l'objectif a été de produire, visualiser et identifier la protéine d'intérêt. La caractérisation de la fonction enzymatique n'a été envisagée que dans un second temps.

Deux principales constructions avec l'insert *IgTeCe* natif ont été testées, *pET28-IgTeCe* et *pET52-IgTeCe*. La souche Rosetta™ 2 a par la suite été transformées avec ces deux constructions. Les premiers paramètres de culture testés ont été largement décrits dans la littérature : une température de 37 °C, une concentration d'1 mM d'IPTG dans le milieu de culture et un temps d'induction court (Das *et al.*, 2011; Upadhyay *et al.*, 2012). Une culture sans IPTG a également été testée comme témoin non induit et a servi de référence (Zhang *et al.*, 2011). Dès ce premier test d'expression, l'analyse Western blot a révélé un signal pour un poids moléculaire correspondant à *IgTeCe*. Cependant, l'analyse SDS-PAGE n'a pas permis de visualiser une surexpression. Au vu de ce résultat, l'expression d'*IgTeCe* a été testée en diminuant la température durant l'induction comme indiqué dans la littérature pour tenter d'améliorer la surexpression (Makrides, 1996). Trois températures, 16, 30 et 37 °C ont été testées et comparées pour les systèmes d'expression (Berrow *et al.*, 2006). Les résultats n'ont pas permis d'observer une surexpression quels que soient la température ou le vecteur utilisés. Seule une expression « induite » (par opposition à une expression basale et une surexpression) a pu être montrée. Les résultats obtenus ont toutefois permis de sélectionner le système d'expression *pET52-IgTeCe*/Rosetta™ 2 comme système à privilégier pour les tests d'activités et pour caractériser un transcrit correspondant à *IgTeCe*. Pour ce système, les profils d'électrophorèse ont révélé les conditions d'expression les plus intéressantes aux températures de 16 et 37 °C. Les résultats ont mis en évidence des transcrits, dont la taille correspond à la séquence codante d'*IgTeCe*, pour les deux températures d'expression confirmant le résultat du Western blot. En conclusion, la protéine *IgTeCe* et son transcrit ont été obtenus lors de l'induction de l'expression dans la bactérie. Le Western blot révèle une seconde bande supérieure de 10 kDa à la bande majeure. Notre hypothèse est la présence d'une forme partiellement phosphorylée de la protéine d'intérêt, cette modification post-traductionnelle étant connue dans la bactérie (Mijakovic, 2010). Les autres

bandes révélées par Western blot à des poids moléculaires inférieurs sont supposées être des produits de dégradation.

La protéine d'intérêt *IgTeCe* est présente dans la bactérie mais les différentes stratégies de purification se sont révélées infructueuses. La recherche d'activité a alors été envisagée sur les extraits totaux de protéines. La méthode consiste à tester l'effet de la présence de la protéine *IgTeCe* sur l'activité lipolytique globale de l'extrait protéique. Dans un premier temps, une activité carboxylestérase a été recherchée sur des esters de para-nitrophényle qui s'est révélée négative. Néanmoins, ce type d'expérimentation est plus habituellement réalisé sur des extraits purifiés d'enzymes (Choi *et al.*, 2004). Le résultat peut s'expliquer de deux manières. La première est une trop faible expression d'*IgTeCe* ne permettant pas de mettre en évidence une activité avec ce test. La seconde est que l'activité testée ne correspond pas à l'activité recherchée. L'absence d'activité significative constatée pour un autre substrat des carboxylestérases, la tributyrine, nous a orienté vers la seconde hypothèse. Une recherche d'activité pour la fonction thioestérase a donc été entreprise. Elle a consisté en l'étude du profil en acides gras des lipides totaux de la bactérie en présence d'*IgTeCe*. La littérature montre des exemples de modification du profil d'acides gras dans la bactérie lors de l'expression hétérologue de protéines annotées comme thioestérase. C'est le cas notamment pour l'étude de Zhang *et al.* (2011) qui a montré la variation des profils d'acides gras obtenus en fonction de la thioestérase exprimée.

Les résultats obtenus au cours de ce travail ont mis en évidence, d'une part, une modification du profil d'acides gras en présence de la protéine et, d'autre part, une différence d'activité d'*IgTeCe* entre 16 °C et 37 °C.

Afin de vérifier que ces modifications sont cohérentes avec l'activité d'une thioestérase, nous avons comparé nos résultats à différentes études de caractérisation de thioestérases. Une des études correspondant à notre problématique est l'étude de Zheng *et al.* (2012), qui a utilisé un système d'expression proche de nos conditions expérimentales (*pET/BL21*). Leur approche est complétée par rapport à nos conditions par l'ajout de glucose ou de glycérol à 20 g/L comme source de carbone. Une modification du profil en acides gras a été mise en évidence. Cependant, aucune correspondance n'est observée entre leur profil et le profil obtenu lors de l'expression d'*IgTeCe*.

Comme nous l'avons indiqué lors de l'analyse des profils en acides gras obtenus en présence de la protéine *IgTeCe*, une spécificité semble exister à 16 °C pour le C16 :1 uniquement et à 37 °C pour le C16 :1 et le C18 :1. Nous nous sommes donc intéressés aux études présentant des spécificités

précises de thioestérases. La première étude réalisée par Jha *et al.* (2010) a montré des spécificités pour le C16:0 et le C18:0. Une autre étude déjà évoquée pour la thioestérase PtTE de la microalgue *Phaeocactylum tricornutum* a montré des spécificités significatives pour le C16:1 et C18:1 correspondant à celles observées pour *IgTeCe*. La séquence protéique de la thioestérase PtTE présente de nombreuses homologues avec les thioestérases FatA de plante présentant une conformation structurale de type « Hotdog ». Cette dernière étude a permis de nous intéresser à la famille d'enzyme FatA et particulièrement à l'étude de l'Acyl-ACP thioestérase FatA issue de l'arachide (*Arachis hypogaea*). L'expression hétérologue de cette Acyl-ACP thioestérase a montré un profil d'acides gras très similaire à celui observé pour *IgTeCe* (Chen *et al.*, 2012).

En conclusion, les résultats observés dans le changement du profil en acides gras des lipides totaux de la bactérie correspondent à ceux observés lors de l'expression de gènes de thioestérases de la famille FatA. Néanmoins, les thioestérases de la famille FatA ne présentent pas une conformation d'hydrolase α/β mais une conformation « Hotdog » (Cantu *et al.*, 2010). Le diagramme en figure 62 permet de localiser à quel stade intervient une thioestérase exprimée de manière hétérologue dans la bactérie (Zhang *et al.*, 2011). La thioestérase *IgTeCe* serait active au niveau du cycle de synthèse des acides gras dans la bactérie.

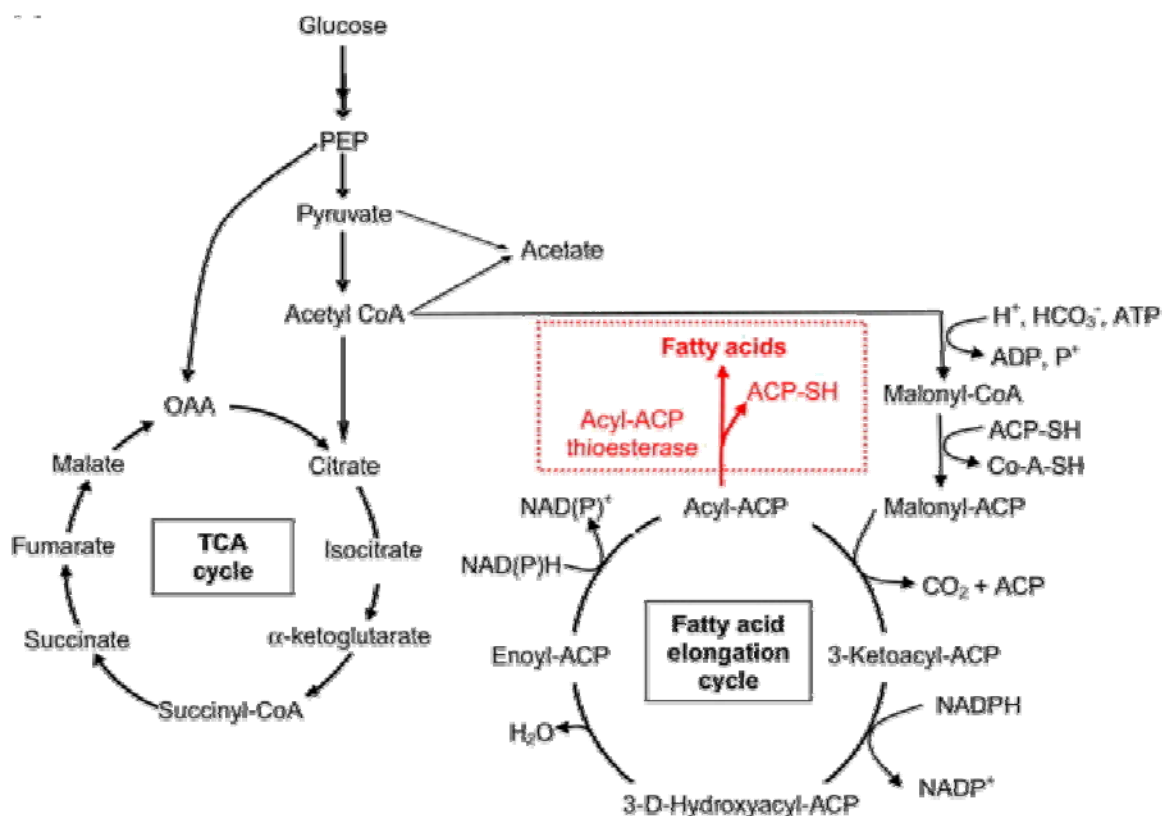


Figure 62: Vue simplifiée de la synthèse d'acides gras chez la bactérie *Escherichia coli* (Zhang *et al.*, 2011)

Au vu des informations obtenues jusqu'à présent, la fonction de la protéine *IgTeCe* est compatible avec la fonction d'Acyl-ACP thioestérase. En effet, ce type enzymatique existe également avec la conformation d'hydrolase α/β (Cantu *et al.*, 2010). Dans ce contexte, une des prédictions *in silico* serait donc vérifiée et cohérente.

Ce résultat soulève une question concernant la localisation d'*IgTeCe* dans la microalgue *Isochrysis galbana*. En effet, dans une microalgue, la fonction thioestérase est généralement présente au niveau du chloroplaste et annotée FAT (Fatty Acyl Thioestérase) (Radakovits *et al.*, 2010) (Figure 63). Jusqu'à présent, nos recherches *in silico* pour *IgTeCe* ne se sont pas intéressées à sa localisation dans la microalgue. Comme il a été évoqué dans la prédiction *in silico*, aucune méthode généraliste n'a permis de détecter un signal peptidique. Néanmoins, le domaine N-terminal demeure non caractérisé sur la séquence d'*IgTeCe*. La détermination de la localisation de la protéine *IgTeCe* dans la microalgue pourrait permettre d'appuyer l'hypothèse d'une fonction thioestérase.

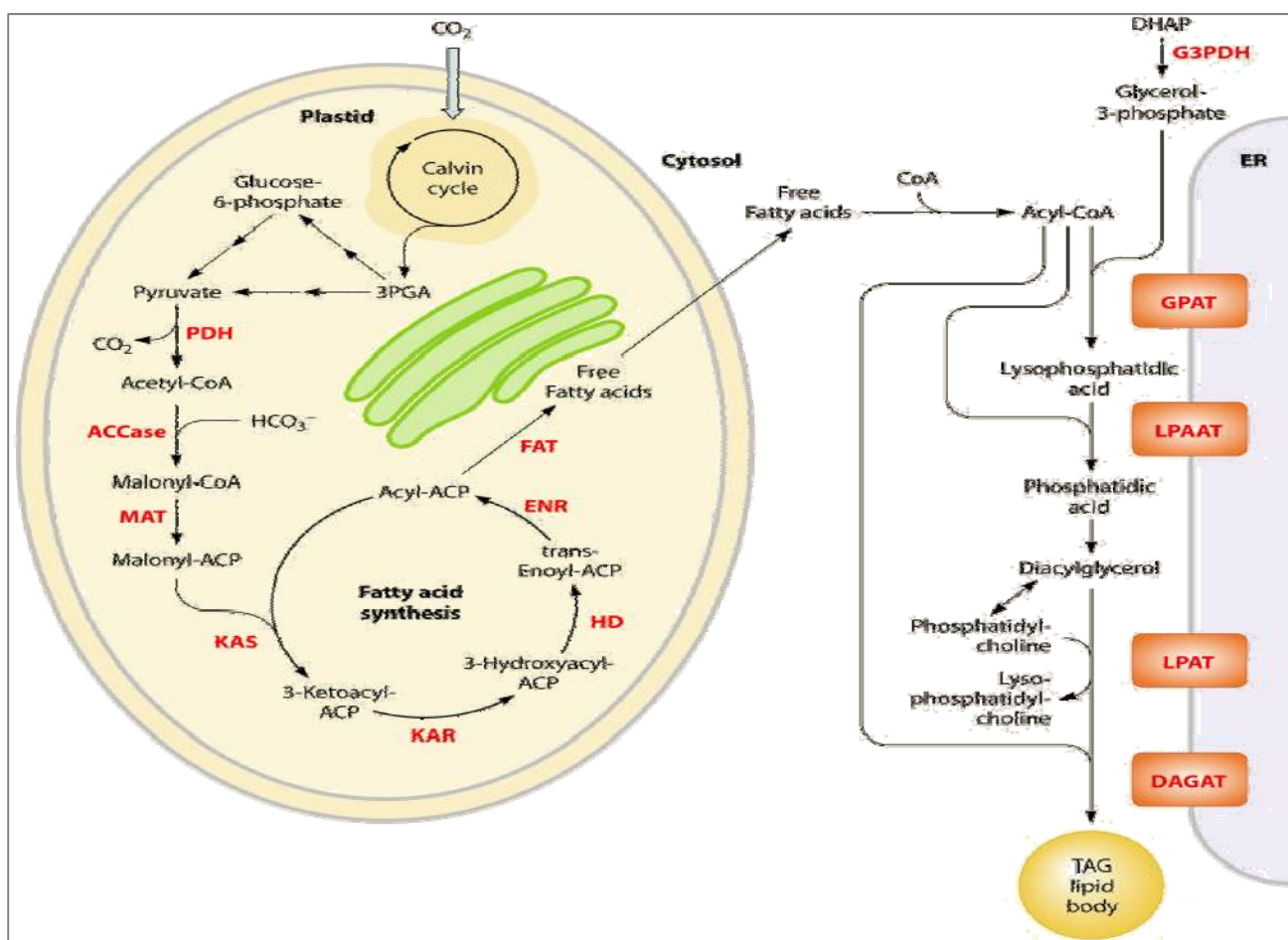


Figure 63: Vue simplifiée des compartiments intervenant dans la synthèse des lipides chez les microalgues (Radakovits *et al.*, 2010)

Pour déterminer cette localisation, des logiciels de localisation chloroplastique existant pour les protéines végétales, ChloroP et PLCR, ont tout d'abord été utilisés (Emanuelsson *et al.*, 1999;

Schein *et al.*, 2001). Les résultats de ces algorithmes prédisent une localisation de la protéine *IgTeCe* dans le compartiment chloroplastique.

Un autre algorithme, PredAlgo, dédié à la localisation de protéines dans les algues, a par la suite été utilisé (Tardif *et al.*, 2012). Cet algorithme analyse la séquence protéique et calcule trois valeurs correspondant à trois localisations cellulaires possibles : mitochondriale, chloroplastique ou extracellulaire (M, C ou SP). Pour que le résultat soit considéré comme pertinent pour valider une localisation, les valeurs seuils à dépasser sont de 0,42 pour une localisation dans la mitochondrie (M) ; de 0,41 pour une localisation dans le chloroplaste (C); de 0,14 pour une sécrétion extracellulaire (SP). Lorsqu'aucune des valeurs seuil n'est atteinte, la localisation cellulaire est définie comme « autre » (O). Les résultats obtenus pour *IgTeCe* sont présentés dans le tableau 61.

Nom	M	C	SP	Compartiment cellulaire	cTPlength
<i>IgTeCe</i>	0.032	1.768	0.0002379	Chloroplaste	88
<i>IgTeCe</i> C-terminal	0.0001643	0.0034	0.0000153	Autre que chloroplaste, mitochondrie, et voie de sécrétion	
<i>IgTeCe</i> N-terminal	0.032	1.768	0.0002379	Chloroplaste	

Tableau 61: Résultats extraits par le logiciel PredAlgo pour la localisation prédictive de la protéine *IgTeCe*.

La localisation déduite de la protéine *IgTeCe* est au niveau du chloroplaste. A partir de cette prédiction, une information complémentaire, « cTPlength », a été extraite et correspond au nombre d'acides aminés théoriquement impliqués dans le domaine d'adressage. Dans le cas présent, le domaine d'adressage d'*IgTeCe* est évalué à 88 acides aminés. L'emplacement de ce domaine n'est cependant pas déterminé par PredAlgo.

Afin de déterminer l'emplacement de ce domaine d'adressage, il a fallu compiler les résultats obtenus par les alignements de séquences BLASTP et le résultat obtenu à l'aide de PredAlgo.

Le domaine hydrolase α/β a été localisé à l'extrémité C-terminal de la protéine. A l'extrémité N-terminal, un domaine d'environ 125 acides aminés est présent mais demeure non identifié par BLASTP. Nous avons donc supposé que ce domaine contient le domaine d'adressage. Pour vérifier cette hypothèse, les 88 acides aminés en N-terminal et les 293 acides aminés en C-terminal ont été analysés séparément par PredAlgo. Le résultat montre que la partie N-terminal est bien celle correspondant au site d'adressage pour le chloroplaste. L'analyse de la partie C-terminal seule révèle une localisation cellulaire « autre ». La localisation cellulaire d'*IgTeCe*, serait donc dépendante du domaine N-terminal.

Au vu de ces résultats, l'utilisation de PredAlgo peut permettre de réorienter l'étude d'*IgEst3*, qui présente également un domaine non structuré en N-terminal. Une analyse semblable à *IgTeCe* a donc été réalisée. Le tableau 62 regroupe les données extraites. La localisation prédite par le logiciel est identique à celle d'*IgTeCe*, soit au niveau du chloroplaste. Le score obtenu pour cette localisation est inférieur à celui d'*IgTeCe*, tout comme la longueur du fragment consacré à cette localisation. Au vu de la conformation de la protéine et de la présence, comme dans le cas d'*IgTeCe*, d'un pentapeptide GX SXG, il peut être supposé que la problématique fonctionnelle d'*IgEst3* puisse être identique à celle que l'on a retrouvée chez *IgTeCe*.

<i>Nom</i>	<i>M</i>	<i>C</i>	<i>SP</i>	<i>Compartiment cellulaire</i>	<i>cTPlength</i>
<i>IgEst3</i>	0.0129885	1.3566809	0.0000838	Chloroplaste	19.59732

Tableau 62: Résultats extraits par le logiciel PredAlgo pour la localisation prédictive de la protéine *IgEst3*.

Afin de confirmer cette démarche, les séquences de deux autres protéines étudiées antérieurement au laboratoire *IgEst2*, une carboxylestérase et *IgLip*, une enzyme similaire à une lipase, ont également été analysées par PredAlgo.

Etant donné leur fonction enzymatique, l'hypothèse de départ est de localiser ces deux enzymes au niveau du cytoplasme de la microalgue et d'invalider leur présence au niveau de la mitochondrie, du chloroplaste ou de l'environnement extracellulaire. Les résultats de cette analyse par PredAlgo sont présentés dans le tableau 63.

<i>Nom</i>	<i>M</i>	<i>C</i>	<i>SP</i>	<i>Compartiment cellulaire</i>
<i>IgEst2</i>	0.0199812	0.1101575	0.5201455	Extracellulaire
<i>IgLip</i>	0.0018817	0.0016639	0.0004140	Autre que chloroplaste, mitochondrie, et voie de sécrétion

Tableau 63: Prédiction pour la localisation cellulaire d'*IgLip* et d'*IgEst2*.

Dans le cas d'*IgEst2*, une orientation de la protéine vers le processus de sécrétion (SP) a été mise en évidence. Dans le cas d'*IgLip*, la prédiction est une orientation vers un compartiment « autre », différent de la mitochondrie ou du chloroplaste. Dans les deux cas, la prédiction de localisation reste compatible avec les fonctions enzymatiques supposées.

CONCLUSION ET PERSPECTIVES

CONCLUSION ET PERSPECTIVES

Ces travaux de thèse ont permis d'amorcer la caractérisation de deux gènes issus de la microalgue *Isochrysis galbana*. Le premier gène code une protéine de 329 acides aminés et de poids moléculaire théorique de 35,41 kDa nommée *IgEst3*. Le second code une protéine de 381 acides aminés et de poids moléculaire théorique de 42,31 kDa nommée *IgTeCe*.

Les fonctions associées à ces séquences protéiques ont été prédites en recherchant des séquences homologues dans plusieurs banques de données et en identifiant leur appartenance à des familles d'enzymes lipolytiques. Les deux séquences protéiques possèdent le pentapeptide consensus Gly-X-Ser-X-Gly et la triade catalytique Ser/Asp/His et peuvent donc être classées dans la famille des hydrolases à repliement α/β .

La fonction prédite de la protéine *IgEst3* est associée à deux familles d'enzymes, thioestérase et carboxylestérase. Elle présente un grand nombre de similitudes avec des enzymes procaryotes. L'approche *in silico* n'a pas permis de favoriser une fonction plutôt qu'une autre. L'expression hétérologue dans la levure s'est révélée infructueuse pour l'obtention de la protéine *IgEst3* empêchant toute caractérisation biochimique.

Le second gène, *IgTeCe*, code une séquence protéique dont la fonction prédite n'a pas été clairement précisée lors des analyses *in silico*. L'expression hétérologue dans la levure a été entreprise pour caractériser la fonction enzymatique. Cependant, il n'a pas été, pour l'instant, possible de détecter ou de purifier la protéine *IgTeCe*.

Le travail a été poursuivi pour *IgTeCe* par expression hétérologue dans la bactérie *Escherichia coli*. Cette démarche a permis d'obtenir la protéine *IgTeCe* recombinante. La présence du transcrit a également été validée. Cependant, les conditions d'expression n'ont pas permis d'obtenir la surproduction attendue. Dans ces conditions, la purification de la protéine *IgTeCe* n'a pas été satisfaisante. Toutefois, l'expression d'*IgTeCe* a permis l'observation d'un changement du profil d'acides gras des lipides totaux de la bactérie. Ce résultat permet, au regard de la littérature, de supposer que la protéine *IgTeCe* a une activité thioestérase *in vivo* chez *Escherichia coli* (Gong *et al.*, 2011).

Les travaux menés au cours de cette thèse ont permis de mettre en évidence l'activité enzymatique de la protéine *IgTeCe*. La recherche de fonction de la protéine a été initiée afin d'identifier les enzymes

responsables de l'activité lipolytique endogène de la microalgue. La protéine *IgTeCe* faisait partie des séquences présentant les caractéristiques d'enzymes lipolytiques recherchées et pouvant être impliquées dans le métabolisme des lipides dans la microalgue. Néanmoins, la fonction Acyl-ACP thioestérase, observée pour la protéine *IgTeCe* lors de son expression hétérologue, aurait bien un rôle d'hydrolyse mais interviendrait plutôt dans le cycle de synthèse des acides gras d'*Isochrysis galbana*.

Pour compléter l'expression dans *Saccharomyces cerevisiae*, il serait intéressant de cloner les séquences d'*IgTeCe* et d'*IgEst3* sans intégrer d'alanine et en testant d'autres vecteurs. De plus, d'autres conditions d'expression sont à tester dans *Saccharomyces cerevisiae*. Enfin, au vu de l'organisation des domaines de chacune des protéines, il pourrait être envisagé de cloner des versions tronquées des séquences codantes.

L'étude de la protéine *IgTeCe* devra être poursuivie. Le bon fonctionnement de la transcription des inserts avec le système d'expression *pET52b+*/Rosetta™ 2 dans la bactérie a été démontré au cours de cette étude. Cependant, il est encore à améliorer, notamment de manière à pouvoir obtenir une surexpression de la protéine d'intérêt. Plusieurs pistes d'étude sont envisageables. Une première piste pourrait être de cloner à nouveau *IgTeCe* en optimisant la distance entre le codon initiateur et le site de fixation du ribosome. Une seconde piste pourrait faire appel à de nouveaux systèmes d'expression (changement de vecteurs et de souches d'expression). Lorsqu'un système fonctionnel de surexpression dans la bactérie aura été trouvé pour *IgTeCe*, il pourra être envisagé d'appliquer le modèle à *IgEst3*.

L'étude de l'activité enzymatique *in vivo* a été initiée au cours de cette étude pour *IgTeCe* et nécessite d'être poursuivie. D'une part, il serait intéressant de compléter les résultats obtenus et de moduler la température d'expression qui semble être un facteur important. En effet, les résultats obtenus à 16 °C comparativement à ceux obtenus à 37 °C montrent un changement d'activité d'*IgTeCe* qu'il serait intéressant d'approfondir. D'autre part, le laboratoire envisage également de tester d'autres sources de carbone (par exemple des sucres) pour observer l'impact sur le changement du profil lipidique. Il peut être également envisagé de créer des versions tronquées de la protéine. Deux domaines composent la protéine et il peut être pertinent de comparer l'activité enzymatique en présence et absence du domaine d'adressage. Cette stratégie pourra également être appliquée pour caractériser biochimiquement les acides aminés qui interviennent dans le site catalytique de la protéine.

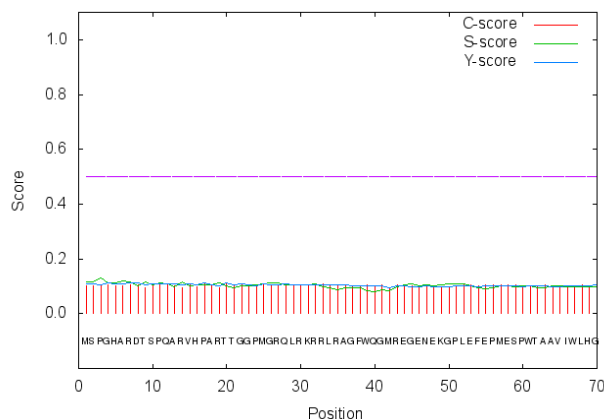
Pour finir, la prédiction de la localisation d'*IgTeCe* dans le chloroplaste de la microalgue a permis de conforter l'hypothèse selon laquelle il s'agirait une thioestérase. La localisation dans un compartiment cellulaire spécifique peut permettre d'orienter la fonction de la protéine. Si cette approche est développée, il peut être envisagé de créer des anticorps spécifiques des protéines *IgTeCe* et *IgEst3*. La prédiction *in silico* de leur localisation pourrait alors être confirmée *in vivo*.

Les travaux menés ont permis la caractérisation d'une enzyme, provenant d'*Isochrysis galbana*, capable de modifier le profil en acides gras au sein des lipides totaux de la bactérie *E. coli*. Cette étude a donc mis en évidence une enzyme impliquée dans le métabolisme lipidique de la microalgue. Cette approche vient compléter les connaissances de la synthèse des acides gras et permet une meilleure compréhension des mécanismes au sein de la microalgue. Il pourra donc être envisagé de continuer l'étude des autres gènes identifiés dans *Isochrysis galbana*.

ANNEXES

ANNEXES

Annexe 1: Sortie graphique de l'outil de prédiction SignalP pour la séquence traduite *IgEst3*. La valeur significative à atteindre est symbolisée par une ligne violette.



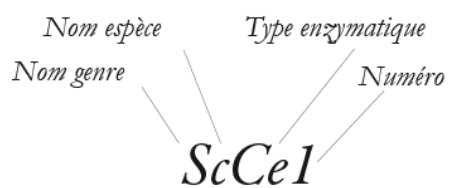
Annexe 2 : Prédiction des sites de N-glycosylation pour la protéine *IgEst3*

Position	Résidu	Score	Prédiction
47	NEK	-1.0021666	Non-glycosylé
136	NNF	-1.1971862	Non-glycosylé
137	NFA	-1.0533533	Non-glycosylé
157	NLP	-1.2381745	Non-glycosylé
181	NYP	-0.96383235	Non-glycosylé
263	NAA	-1.2373241	Non-glycosylé
292	NPA	-1.5907099	Non-glycosylé

Annexe 3 : Prédiction des sites d'O-glycosylation pour la protéine *IgEst3*

<i>IgEst3</i>							
NetOGlyc				GlycoEP			
Numéro de résidu	Score	Acide aminé	Résultat	Numéro de résidu	Score	Acide aminé	Résultat
2	0.820146	S	Potentiellement glycosylé	2	0.91243996	S	Potentiellement glycosylé
9	0.86302	T	Potentiellement glycosylé	9	0.093605695	T	Potentiellement glycosylé
10	0.945968	S	Potentiellement glycosylé	10	0.51630892	S	Potentiellement glycosylé
20	0.98983	T	Potentiellement glycosylé	21	0.55874728	T	Potentiellement glycosylé
21	0.988343	T	Potentiellement glycosylé	91	0.029994059	S	Potentiellement glycosylé
200	0.515674	S	Potentiellement glycosylé	113	0.031753327	T	Potentiellement glycosylé
203	0.5	T	Potentiellement glycosylé	133	0.26941402	S	Potentiellement glycosylé
205	0.601515	S	Potentiellement glycosylé	134	0.14576111	S	Potentiellement glycosylé
209	0.550038	S	Potentiellement glycosylé	141	0.18970636	T	Potentiellement glycosylé
211	0.732257	S	Potentiellement glycosylé	169	0.35393795	S	Potentiellement glycosylé
289	0.852496	T	Potentiellement glycosylé	200	0.17781876	S	Potentiellement glycosylé
301	0.865845	T	Potentiellement glycosylé	205	0.57891986	S	Potentiellement glycosylé
302	0.69926	S	Potentiellement glycosylé	209	0.16181966	S	Potentiellement glycosylé
307	0.731433	S	Potentiellement glycosylé	211	0.4576465	S	Potentiellement glycosylé
318	0.587081	S	Potentiellement glycosylé	221	0.35661475	S	Potentiellement glycosylé
				224	0.20242807	T	Potentiellement glycosylé
				238	0.23667494	T	Potentiellement glycosylé
				244	0.017319906	S	Potentiellement glycosylé
				289	0.13118333	T	Potentiellement glycosylé
				301	0.062438994	T	Potentiellement glycosylé
				307	0.15489306	S	Potentiellement glycosylé

Annexe 4 : Annotation des séquences alignées



Ce = Carboxylestérase

Lp = Lysophospholipase

Pl = Phospholipase

Te = Thioestérase

PlCe = Phospholipase/ Carboxylestérase

Est = Ester Hydrolase

TeCe = Thioestérase/ Carboxylestérase

Hp = Protéine hypothétique

Upp = Protéine inconnue

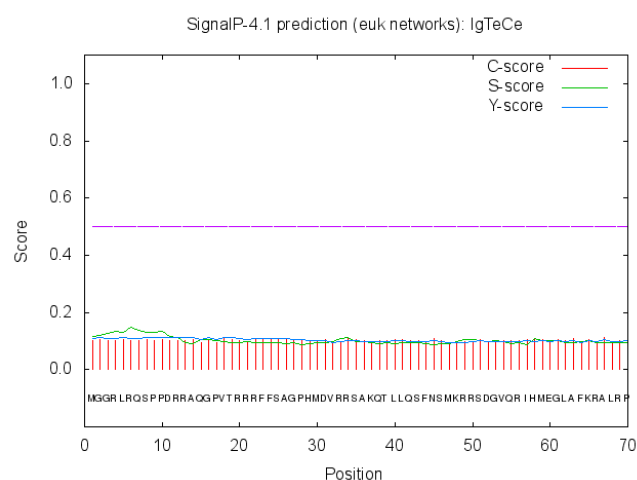
Annexe 5 : Numéros d'accès des séquences alignées à *IgEst3* et *IgTeCe*.

Tableau 1

Annotations	Nom complet	Homologue avec	Numéro d'accès
 AlCe1 	<i>Arabidopsis lyrata</i> Carboxylestérase	<i>IgEst3</i>	ref XP_002867777.1
 ApTe1_2 	<i>Acyrtosiphon pisum</i> Thioestérase like	<i>IgEst3</i>	ref NP_001155486.1
 AtCe2 	<i>Arabidopsis thaliana</i> Carboxylestérase 2	<i>IgEst3</i>	NP_001190797.1
 AtPlCe2 	<i>Arabidopsis thaliana</i> Phospholipase/Carboxylestérase 2	<i>IgEst3</i>	NP_197506
 AtSOBER1 	<i>Arabidopsis thaliana</i>	<i>IgEst3</i>	gb ABO26813.1
 BbCe1 	<i>Bordetella bronchiseptica</i> Carboxylestérase	<i>IgEst3</i>	gi 33603335
 BpCe1 	<i>Bathycoccus prasinos</i> Carboxylestérase	<i>IgEst3 / IgTeCe</i>	emb CCO15175.1
 CaTe1 	<i>Candida albicans</i> Thioestérase	<i>IgEst3</i>	sp Q5AGD1.2
 CcCe1 	<i>Chondrus crispus</i> Carboxylestérase	<i>IgEst3</i>	ref XP_005714993.1
 DhTe1 	<i>Debaryomyces hansenii</i> Thioestérase	<i>IgEst3</i>	sp Q6BSS8.2
 EhHp1 	<i>Emiliana huxleyi</i> Hypothetical protein 1	<i>IgEst3</i>	XP_005790747
 FiCe1 	<i>Francisella tularensis</i> Carboxylestérase	<i>IgEst3 / IgTeCe</i>	PDB: 4F21_A
 FvTe1 	<i>Fragaria vesca</i> Thioestérase 1	<i>IgEst3</i>	XP_004293347.1
 FvTe2 	<i>Fragaria vesca</i> Thioestérase 2	<i>IgEst3</i>	XP_004293348.1
 GcTe1 	<i>Glaciecola chathamensis</i> Thioestérase	<i>IgEst3 / IgTeCe</i>	ref WP_007984303.1
 GtHp1 	<i>Guillardia theta</i> Hypothetical protein	<i>IgEst3</i>	gi 551651019
 HeCe1 	<i>Herbaspirillum</i> sp. Carboxylestérase	<i>IgEst3</i>	ref WP_008327795.1
 HfCe1 	<i>Herbaspirillum frisingense</i> Carboxylestérase	<i>IgEst3</i>	ref WP_006463578.1
 HsTe1 	<i>Homo sapiens</i> Thioestérase	<i>IgEst3 / IgTeCe</i>	PDB : 1FJ2_A
 IdCe1 	<i>Idiomarina</i> sp. Carboxylestérase	<i>IgEst3</i>	ref WP_007421277.1
 JcPlCe1 	<i>Jatropha curcas</i> Phospholipase/Carboxylestérase 2	<i>IgEst3 / IgTeCe</i>	dbj BAJ53250.1
 KkCe1 	<i>Kangiella koreensis</i> Carboxylestérase	<i>IgEst3</i>	ref YP_003145361.1
 NgLp2 	<i>Nannochloropsis gaditana</i> Lysophospholipase2	<i>IgEst3/IgTeCe</i>	XP_005853662.1
 NgLp3 	<i>Nannochloropsis gaditana</i> Lysophospholipase like2	<i>IgEst3</i>	gb EKU22703.1
 OcLp1 	<i>Oryctolagus cuniculus</i> Lysophospholipase	<i>IgEst3</i>	gi 41017275
 OdUpp1 	<i>Oikopleura dioica</i> Unknown protein product	<i>IgEst3</i>	emb CBY40901.1
 OsLp1 	<i>Ophiocordyceps sinensis</i> Lysophospholipase	<i>IgEst3</i>	gb EQL00530.1
 OsLp2 	<i>Oryza sativa</i> Lysophospholipase2	<i>IgEst3</i>	Q5ZBI5
 OtLp1 	<i>Ostreococcus tauri</i> Lysophospholipase	<i>IgEst3 / IgTeCe</i>	ref XP_003080549.1
 PfCe2S 	<i>Pseudomonas fluorescens</i> Carboxylestérase 2	<i>IgEst3 / IgTeCe</i>	PDB : 1AUO_A
 PsCe1 	<i>Pseudomonas synxantha</i> Carboxylestérase	<i>IgEst3</i>	ref WP_005785043.1
 RcOSBS 	<i>Ricinus communis</i>	<i>IgEst3</i>	ref XP_002514555.1
 RcTe1 	<i>Ricinus communis</i> Thioestérase	<i>IgEst3</i>	ref XP_002517576.1
 RfCe1 	<i>Rhodospirillum rubrum</i> Carboxylestérase	<i>IgEst3</i>	ref YP_522935.1
 RnLp1 	<i>Rattus norvegicus</i> Lysophospholipase	<i>IgEst3</i>	sp P70470.1
 TaCe1 	<i>Taylorella asinigenitalis</i> Carboxylestérase	<i>IgEst3</i>	ref YP_007821969.1
 TrHp1 	<i>Trichoderma reesei</i> Hypothetical protein	<i>IgEst3</i>	gb EGR46897.1
 TrPp1 	<i>Trichoderma reesei</i> Predicted protein	<i>IgEst3</i>	gb EGR46897.1
 UbCe1 	Uncultured bacterium Carboxylestérase	<i>IgEst3</i>	gb ACZ16579.1
 UmTe1 	<i>Ustilago maydis</i> Thioestérase	<i>IgEst3</i>	sp Q4PID3.1
 VvTe1 	<i>Vitis vinifera</i> Thioestérase 2	<i>IgEst3</i>	ref XP_003635387.1
 ZmTe1 	<i>Zea mays</i> Thioestérase 1	<i>IgEst3</i>	ref NP_001152004.1

Tableau2

Annotations	Nom complet	Homologue avec	Numéro d'accès
AaHp1	<i>Aureococcus anophagefferens</i> Hypothetical protein	IgTeCe	gb EGB07635.1
BmTe2	<i>Bos mutus</i> Thioestérase 2	IgTeCe	ELR49448.1
CeATH1	<i>Caenorhabditis elegans</i> Thioestérase	IgTeCe	Q21224_CAEEEL
CmLp2	<i>Cyanidioschyzon merolae</i> Lysophospholipase2	IgTeCe	ref XP_005535813.1
CsCe1	<i>Chromohalobacter Saalexigens</i> Carboxylestérase	IgTeCe	ref YP_573162.1
CsPlCe1	<i>Coccomyxa subellipsoidea</i> Phospholipase/Carboxylestérase	IgTeCe	ref XP_005649811.1
DdTe1h2	<i>Dictyostelium discoideum</i> Thioestérase homologue	IgTeCe	sp Q55FK4.1
EhHp2	<i>Emiliana huxleyi</i> Hypothetical protein 2	IgTeCe	Ref XP_005775234
EhHp3	<i>Emiliana huxleyi</i> Hypothetical protein 3	IgTeCe	ref XP_005793414.1
EsLp2	<i>Ectocarpus siliculosus</i> Lysophospholipase2	IgTeCe	emb CBJ31154.1
GaTe1	<i>Glaciecola agarilytica</i> Thioesterase	IgTeCe	ref WP_008305204.1
GnPlCe1	<i>Glaciecola nitratreducens</i> Phospholipase/Carboxylestérase	IgTeCe	YP_004870356
GpTe1	<i>Glaciecola pallidula</i> Thioestérase	IgTeCe	ref WP_006008301.1
GsLp2	<i>Galdiera sulphuraria</i> Lysophospholipase2	IgTeCe	ref XP_005703377.1
GzTe1	<i>Gibberella zeae</i> Thioestérase	IgTeCe	sp Q4I8Q4.1
HsLYPL1S	<i>Homo sapiens</i> Lysophospholipase	IgTeCe	pdb 3U0V A
MgCe1	<i>Methylovorus glucosetrophus</i> Carboxylestérase	IgTeCe	ref YP_003050375.1
MmTe2	<i>Mus musculus</i> Thioestérase 2	IgTeCe	ref NP_036072.1
NgLp1	<i>Nannochloropsis gaditana</i> Lysophospholipase like1	IgTeCe	NW_005803674.1
PaCe1	<i>Pseudomonas aeruginosa</i> Carboxylestérase	IgTeCe	pdb 3CN7 A
PfCe2S	<i>Pseudomonas fluorescens</i> Carboxylestérase 2	IgTeCe	pdb 1AUO
RcTe2	<i>Ricinus communis</i> Thioestérase 2	IgTeCe	ref XP_002524444.1
RnTe2	<i>Rattus norvegicus</i> Thioestérase 2	IgTeCe	NP_112632.1
SaCe1	<i>Shewanella amazonensis</i> Carboxylestérase	IgTeCe	YP_927897
ScpTe1	<i>Schizosaccharomyces pombe</i> Thioestérase	IgTeCe	sp O42881.1
TcEst1	<i>Theobroma cacao</i> Ester hydrolase	IgTeCe	ref XP_007013352.1
TpPl1	<i>Thalassiosira pseudonana</i> Phospholipase	IgTeCe	ref XP_002289178.1

Annexe 6 : Sortie graphique de l'outil de prédiction SignalP pour la protéine *IgTeCe*Annexe 7 : Prédiction des sites de N-glycosylation pour la protéine *IgTeCe*

Position	Résidu	Score	Prédiction
45	N S M	-0.83124062	Non-glycosylé
198	N V A	-1.0992968	Non-glycosylé
204	N Y Y	-0.94181693	Non-glycosylé
219	N E A	-1.0234075	Non-glycosylé
266	N Y H	-1.3226977	Non-glycosylé

Annexe 8 : Prédiction des sites d'O-glycosylation pour la protéine *IgTeCe*

IgTeCe							
NetOGlyc				GlycoEP			
Numéro de résidu	Score	Acide aminé	Résultat	Numéro de résidu	Score	Acide aminé	Résultat
8	0.964061	S	Potentiellement glycosylé	19	0.19772839	T	Potentiellement glycosylé
19	0.88193	T	Potentiellement glycosylé	78	0.5987724	S	Potentiellement glycosylé
25	0.836416	S	Potentiellement glycosylé	83	0.62931816	S	Potentiellement glycosylé
35	0.555844	S	Potentiellement glycosylé	84	0.36119261	S	Potentiellement glycosylé
43	0.507754	S	Potentiellement glycosylé	86	0.32615738	S	Potentiellement glycosylé
74	0.672293	S	Potentiellement glycosylé	88	0.108852	S	Potentiellement glycosylé
75	0.897631	S	Potentiellement glycosylé	89	0.80962281	S	Potentiellement glycosylé
78	0.816967	S	Potentiellement glycosylé	91	0.28938989	S	Potentiellement glycosylé
83	0.670997	S	Potentiellement glycosylé	99	0.12967567	T	Potentiellement glycosylé
84	0.93868	S	Potentiellement glycosylé	106	0.50515594	S	Potentiellement glycosylé
86	0.797425	S	Potentiellement glycosylé	109	0.079237434	S	Potentiellement glycosylé
88	0.684262	S	Potentiellement glycosylé	119	0.31754896	S	Potentiellement glycosylé
89	0.912344	S	Potentiellement glycosylé	125	0.067479208	T	Potentiellement glycosylé
91	0.517386	S	Potentiellement glycosylé	126	0.35442927	T	Potentiellement glycosylé
119	0.854023	S	Potentiellement glycosylé	127	0.46072118	S	Potentiellement glycosylé
125	0.761641	T	Potentiellement glycosylé	144	0.19884173	T	Potentiellement glycosylé
126	0.828983	T	Potentiellement glycosylé	158	0.060157845	S	Potentiellement glycosylé
127	0.854691	S	Potentiellement glycosylé	190	0.23886172	S	Potentiellement glycosylé
207	0.5493	T	Potentiellement glycosylé	193	0.16971214	T	Potentiellement glycosylé
225	0.506306	S	Potentiellement glycosylé	254	0.12547797	S	Potentiellement glycosylé
293	0.710352	S	Potentiellement glycosylé	264	0.093693359	T	Potentiellement glycosylé
358	0.558794	T	Potentiellement glycosylé	279	0.092311497	S	Potentiellement glycosylé
370	0.578553	S	Potentiellement glycosylé	286	0.60584428	T	Potentiellement glycosylé
				344	0.019048772	S	Potentiellement glycosylé

Annexe 9 : Séquençages des plasmides *pJET-IgTeCe**pJET-Ex1-partie1*

Alignement de la séquence théorique *pJET-IgTeCe-Ex1* (seq_1) avec la séquence obtenue par séquençage (seq_2).

```

Seq_1 241 ttcacatagtggaataaaatcaactgctttaacacttgctgacctgaacaccatatccatcc 300
Seq_2 10 ----- 9

Seq_1 301 ggcgtaatacgaactcactatagggagagcgcgccagatcttcggatggctcgagttt 360
Seq_2 10 -----gatggctcgAGTTT 23
                        XhoI

Seq_1 361 ttcagcaagatatggcgccgctcttcgtcaaagccctccagaccggcgcccgaggggc 420
Seq_2 24 TTCAGCAAGATATGGCGGCCGCTCTTCGTCAAAGCCCTCCAGACCGCGCGCCAGGGGC 83
                        Début d'IgTeCe
                        dans pJET1.2 Ex1.

Seq_1 421 cggcgactcgccgcgctttcttctcagccggcccccacatggacgtccggcggtccgcca 480
Seq_2 84 CGGTGACTCGCGCGCTTCTTCTCAGCCGGCCCCACATGGACGTCCGGCGGTCCGCCA 143

Seq_1 481 agcagaccctcttcagtccttcaatagcatgaagcgctgcagcgacggcggtccaacgta 540
Seq_2 144 AGCAGACCTTCTTCAGTCCTTCAATAGCATGAAGCGTCGCAGCGACGGCGTCCAACGTA 203

Seq_1 541 tccacatggaggggctggcattcaaacgagcgctccggcctcgaggatttcgtcggaag 600
Seq_2 204 TCCACATGGAGGGGCTGGCATTCAAACGAGCGCTCCGGCCTCGCAGGTATTCTGTCGGACG 263

Seq_1 601 catccagaatcgccgcttccccaagtcgcaatcctcgctctcggaaggcaagaggaca 660
Seq_2 264 CATCCAGAATCGCCGCTTCTCCAAGTCGCAATCCTCGCTCTCGGACGCAAGAGGACA 323

Seq_1 661 gcgagaccgatttgatgacgactgcagcgatgcctcgatgatcgacggatatctggcgg 720
Seq_2 324 GCGAGACCGATTGATGACGACTGCAGCGATGCCTCGATGATCGACGGATATCTGGCGG 383

Seq_1 721 cccctcaggtcggattcgcgccacgacgtccgaggatcccggtatcgtggtcatcgagc 780
Seq_2 384 CCCCTCAGGTCGGATTGCGCGCACGACGTCCGAGGATCCCGGTATCGTGGTCATCGAGC 443

Seq_1 781 cctctgggtccacacgcacacgcacccattctattgcacggcatgtactgtctccggaga 840
Seq_2 444 CCTCTGGTCCACACACGCACACCGCCATTCTATTGCACGGCATGTACTGCTCTCCGGAGA 503

Seq_1 841 gctccgatacaattgtcggtctacgtgtgtgtaaatcccttggcctgctgccaggga 900
Seq_2 504 GCTCCGATACAATTGTCGGTCTACCTGCTGCTGTAATCCCTTGGCCTGCTGCCAGGGA 563

```

pJET-Ex1-partie 1

Seq_1	901	ttaaatatgtctctcccgcatgcgcgcgctcggaaccatcagctggccgacagccccgagg	960
Seq_2	564	TTAAATATGTCTTCCCGCATGCGCCGCGTCGGACCATCAGCTGGCCGACAGGCCCGGAGG	623
Seq_1	961	caaacgttgctcctcgtgtacaattactacacgcgcgcgggatggcgagctcgagcatgacg	1020
Seq_2	624	CAAACGTTGCCTCCTGGTACAATTACTACACGCGCCGGATGGCGAGCTCGAGCATGACG	683
Seq_1	1021	tcttgaatgaggcgcacctggcttcgcaaacccggcggaatccactccattgtcgagcgcg	1080
Seq_2	684	TCTTGAATGAGGCGCACCTGGCTTCGCAAAACCGCGGATCCACTCCATTGTGCGAGCGCG	743
Seq_1	1081	aggctgcgcttcttggggggcagcgcgcgcgcgcacatcgtctcggtgggagctcacaagggg	1140
Seq_2	744	AGGCTGCGCTTCTTGGGGGCGACGCGCGCCGCATCATGCTCGGTGGGAGCTCACAAGGGG	803
Seq_1	1141	gcacagtcgctctccatgctacgatcaattaccatcggcctctcggcgcaactgctttgtc	1200
Seq_2	804	GCACAGTCGCTCTCCATGCTACGATCAATTACCATCGGCCTCTCGGCGCACTGCTTTGTG	863
Seq_1	1201	ttcgctcatgtctcatcgattcgggtcacgctttccgagagacaagcgctcccccgccgcgcg	1260
Seq_2	864	TTCGCTCATGTCTCATCGATTCCGGTCACGTTTCCGAGAGACAAGCGCTCCCCCGCCGCCG	923
Seq_1	1261	gcacacccgctcttcgtctctcgcgcgcgcgcaggacaaggtgtatgccccacagctgcagt	1320
Seq_2	924	GCACACCCGCTCTTGCTCTTCGCCGCCGCCAGGACAAGGTGTATGCCCCACAGCTGCAGT	983
Seq_1	1321	atcgcggcttctctctctgtcgtggcgactctggtcttcacgtggagtggcatgtggaacctc	1380
Seq_2	984	ATCGCGGCTTCTCTCTGTCTGGCCGACTCTGGCTTCCACGTGGAGTGGCATGTGGAACCTC	1043
Seq_1	1381	atctgacgcactgggacgagtcgcgcacagagaaaacgcgtagtcgctgcatggatgcgcgc	1440
Seq_2	1044	ATCTGAGCGCACTGGGACGAGTTCGGCGACAGAGAAACGCGTAGTCGCTGCATGGATCGCGC	1103

pJET-Ex1-partie 2

Seq_1	1258	cgcgcacacccctcttcgtcttcgcgcgcgcgcagcaggaagaagtgtagtccccacagctgc	1317	
Seq_2	280	CGGCGACACCCGCTTCTCGTCTTCGCCGCCGCCAGGACAAAGGTGTATGCCCCACAGCTGC	221	
Seq_1	1318	agtatcgcggcttctctctgctgcccgcactctggcttccacgtggagtggcattgtggaac	1377	
Seq_2	220	AGTATCGCGGCTTCTCTCTGCTGGCGACTCTGGCTTCCACGTGGAGTGGCATGTGGAAC	161	
Seq_1	1378	ctcatctgacgcactgggacgagtgcgcacagagaaaacgcgtagtcgctgcatggatcg	1437	
Seq_2	160	CTCATCTGACGCACTGGGACGAGTCGCGCACAGAGAAAACGCGTAGTCGCTGCATGGATCG	101	
Seq_1	1438	cgcgcacagtccgagagcgcgggctccgagagcgcgagcgctccgtagtcgcctgagctta	1497	Fin de séquençage partie 1
Seq_2	100	CGCGCACAGTCCGAGAGCGCGGGCTCCGAGAGCGCGAGCGCTCCGTATCGCCTGAGCTTA	41	
Seq_1	1498	tcctggagccctcttttcattctttctagaagatctcttacaattattctcagctgccatgga	1557	Fin d'IgTeCe dans pJET1.2 ExL
Seq_2	40	TCCTGGAGCCTCTTTTCATCTTTCTAGAGATctcgtaccaa-----	1	
		XbaI		

pJET-Ex2-partie1

Alignement de la séquence théorique *pJET-IgTeCe-Ex2* (seq_1) avec la séquence obtenue par séquençage (seq_2).

```

Seq_1 1599 -cagcctgaaaatcttgagagaataaaagaagaacatcg-atTTT-cCATGGCAGCTGAG 1543
Seq_2 1203 ccagcctgaaaatcttgagggaataaaagaagaacatCGGATTTTCCCATGGCAGCTGAG 1144

Seq_1 1542 aatattgtaggagatcttctagaagatafgggcggcggtcttctgtcaaagccctccaga 1483
Seq_2 1143 AATATTGTAGGAGATCTTCTAGAAAATATGggGCGGCCGCTCTTCGTCAAAGCCCTCCAGA 1084
                                     XbaI
Seq_1 1482 cggcgcgcccgaggggcggtgactcgccgcggtttcttctcagcgggccccacatgga 1423
Seq_2 1083 CCGGCGCGCCAGGGGCGGTGACTCGCCGCCGTTTCTTTTCAGCGGCCCCACATGGA 1024

Seq_1 1422 cgtcggcggttcggccaagcagacccttcttcagtccttcaatagcatgaagcgtgcag 1363
Seq_2 1023 CGTCGCGGTCGCGCCAAGCAGACCCCTTCTTCAGTCCCTCAATAGCATGAAGCGTCGCAG 964

Seq_1 1362 cgacggcgtccaacgtatccacatggagggtggtgcatcaaacgagcgctccggcctcg 1303
Seq_2 963 CGACGGCGTCCAACGTATCCACATGGAGGGGCTGGCATTCAAACGAGCGCTCCGGCCTCG 904

Seq_1 1302 cagggtattcgtcgagcgcacccagaatcgccgcttcttccaagtgcgaatcctcgctctc 1243
Seq_2 903 CAGGTATTCTGTCGACGCATCCAGAATCGCGCTTCTTCCAAGTCGAATCCTCGCTCTC 844

Seq_1 1242 ggacggcgaagaggcagcgagaccgatttggtgacgactgcagcagatgcctcgatgat 1183
Seq_2 843 GgACGCGAAAGAGGACAGCGAGACCGATTGGATGACGACTGCAGCGATGCCTCGATGAT 784

Seq_1 1182 cgacggatatactggcgcccccctcaggtcggattcggcgacagcgtccagggatccgg 1123
Seq_2 783 CGACGGATATCTGCGGCCCTCAGGTCGGATTGGCGCACGACGTCGAGGATCCCGG 724

Seq_1 1122 tatcgtggtcatcgagccctctggtccacacgcacacccgcatcttatgacggcat 1063
Seq_2 723 TATCGTGGTCATCGAGCCCTCTGCTCCACACACGACACCGCCATTCTATTGACGGCAT 664

Seq_1 1062 gtactgctctcggagagctccgatacaaatgtcggtetacctgctgctgcaaatccct 1003
Seq_2 663 GTACTGCTCTCGGAGAGCTCCGATACAAATGTCGGTCTACTGCTGCTGTCAATCCCT 604

Seq_1 1002 tggcctgctgccaaggattaaatatgtcttcccgcatgcgcgcgctcggaaccatcagctg 943
Seq_2 603 TGGCCTGCTGCCAGGATTAATATGTCTTCCCGCATGCGCGCTCGGAACCATCAGCTG 544

Seq_1 942 gccgacaggccccgaggcaaacgttgctcctggtacaattactacacgcgcgggatgg 883
Seq_2 543 GCCGACAGGCCCGAGGCAACGTTGCTCCTGGTACAATTACTACACGCGCGGGATGG 484

Seq_1 882 cgagctcgagcatgacgtcttgaatgaggcgacacctggcttcgcaaacccggcggaacca 823
Seq_2 483 CGAGCTCGAGCATGACGCTTGAATGAGGCGACCTGGCTTCGCAAAACCGGCGGATCCA 424

Seq_1 822 ctccattgtcgagcgcgaggtctgctcttctggggcgagcgcgcgcatcatgctcgg 763
Seq_2 423 CTCCATTGTGAGCGCGAGGCTGCGCTTCTGGGGCGAGCGCGCGCATCATGCTCGG 364

Seq_1 762 tgggagctcaaacgggggacagtcgctctccatgctaagatcaattaccatcgccctct 703
Seq_2 363 TGGGAGCTCAAGGGGGACAGTCGCTCTCCATGCTACGATCAATTACCATCGGCCTCT 304

```

Début d'IgTeCe dans pJET1.2 Ex2

[illegible]

Annexe 10 : Séquençage du plasmide *pJET-IgTeCe NdeI-NotI*

Partie 1

Alignement de la séquence attendue : *pJET-IgTeCe-NdeI-NotI* avec la séquence obtenue par séquençage *pJET-IgTeCe-NdeI-NotI S*

Similarity : 1137/1151 (98,78 %)

Seq_1	1	atgggcgccgctctctcgtcaaaagccctccagaccggcgccgcccagggcggtgactcgc	60	Début d' <i>IgTeCe-Nde-NotI</i> : le site de restriction n'a pas été intégré.
Seq_2	1	-t-ggcgccgctctctcgtcaaaagccctccagaccggcgccgcccagggcggtgactcgc	58	
Seq_1	61	cgccgctttcttctcagccggcccccacatggacgtccggcggtccgccaagcagaccctt	120	
Seq_2	59	cgccgctttcttctcagccggcccccacatggacgtccggcggtccgccaagcagaccctt	118	
Seq_1	121	cttcagtccttcaatagcatgaagcgtcgcagcgacggcggtccaacgtatccacatggag	180	
Seq_2	119	cttcagtccttcaatagcatgaagcgtcgcagcgacggcggtccaacgtatccacatggag	178	
Seq_1	181	gggctggcattcaaacgagcgctccggcctcgcaggtattcgtcgacgcattccagaatc	240	
Seq_2	179	gggctggcattcaaacgagcgctccggcctcgcaggtattcgtcgacgcattccagaatc	238	
Seq_1	241	gcgcgttctctccaagtgcgaatcctcgtctcggacggcaagaggacagcgagaccgat	300	
Seq_2	239	gcgcgttctctCMARKKCGCRW-cctcgtctcggacggcaagaggacagcgagaccgat	297	
Seq_1	301	ttggatgacgactgcagcgatgcctcgatgatcgacggatatctggcgccccctcaggt	360	
Seq_2	298	ttggatgacgactgcagcgatgcctcgatgatcgacggatatctggcgccccctcaggt	357	
Seq_1	361	cggattcggcgacgacgtccgaggatccgggtatcgtggtcatcgagccctctggtcca	420	
Seq_2	358	cggattcggcgacgacgtccgaggatccgggtatcgtggtcatcgagccctctggtcca	417	
Seq_1	421	cacacgcacacccgcatctatttgcacggcatgtactgctctcggagagctccgataca	480	
Seq_2	418	cacacgcacacccgcatctatttgcacggcatgtactgctctcggagagctccgataca	477	
Seq_1	481	tttgtgggtctacctgctgtgtcaaatcccttggcctgctgccagggattaaatgtgc	540	
Seq_2	478	attgtgggtctacctgctgtgtcaaatcccttggcctgctgccagggattaaatgtgc	537	
Seq_1	541	ttcccgcatgcgcgcgtcggaccatcagctggcgacagggcccgaggcaaacgttgcc	600	
Seq_2	538	ttcccgcatgcgcgcgtcGGACCATCAGCTGGCCGACARG-CMCCAKGMAAAMRWKRS	596	
Seq_1	601	tcctgggtacaattactacacgcgcgggatggcgagctcgagcatgacgtcttgaatgag	660	
Seq_2	597	tcctgggtacaattactacacgcgcgggatggcgagctcgagcatgacgtcttgaatgag	656	
Seq_1	661	gcgacacctggcttcgcaaacccggcggtaccactccattgtcgagcgcgaggctgcgctt	720	
Seq_2	657	gcgacacctggcttcgcaaacccggcggtaccactccattgtcgagcgcgaggctgcgctt	716	
Seq_1	721	cttggggggcagcgcgccgcatcatgctcggtgggagctcacaagggggcacagtgcct	780	
Seq_2	717	cttggggggcagcgcgccgcatcatgctcggtgggagctcacaagggggcacagtgcct	776	
Seq_1	781	ctccatgctacgatcaattaccatcggcctctcggcgcaactgctttgtcttcgctcatgt	840	
Seq_2	777	ctccatgctacgatcaattaccatcggcctctcggcgcaactgctttgtcttcgctcatgt	836	
Seq_1	841	ctcatcgattcgggtcacgttttcgagagacaagcgctccccgcgcgcggcacaccgctc	900	
Seq_2	837	ctcatcgattcgggtcacgttttcgagagacaagcgctccccgcgcgcggcacaccgctc	896	

Partie 2

Seq_1	901	ttcgtcttcgcccgcgcccaggacaaggtgtatgccccacagctgcagtatcgcggttc	960
Seq_2	897	ttcgtcttcgcccgcgcccaggacaaggtgtatgccccacagctgcagtatcgcggttc	956
Seq_1	961	tctctgctggcgcgactctggcttccacgtggagtggcatgtggaacctcatctgacgcac	1020
Seq_2	957	tctctgctggcgcgactctggcttccacgtggagtggcatgtggaacctcatctgacgcac	1016
Seq_1	1021	tgggacgagtcgcgcacagagaaacgcgtagtcgctgcatggatcgcgcacagtcoga	1080
Seq_2	1017	tgggacgagtcgcgcacagagaaacgcgtagtcgctgcatggatcgcgcacagtcoga	1076
Seq_1	1081	gagcgcgggctccgagagcgcgagcgcctccgtatcgctgagcttatcctggagcctctt	1140
Seq_2	1077	gagcgcgggctccgagagcgcgagcgcctccgtatcgctgagcttatcctggagcctctt	1136
Seq_1	1141	ttc-----	1143
Seq_2	1137	ttcgcggcgcgtcaa	1151

Fin d'*IgTeCe-Nde-Not*;
le site Not I a bien été intégré.

Annexe 11 : Séquençage du plasmide *pET28-IgTeCe*

McoI
|

tt gga gag aat tac ctc tag aat aat ttt gtt taa ctt taa gaa gga gat ata cca tgg cag ctg aga ata ttg tag gag atc ttc tag aaa gat atg gg < 100 Début d'*IgTeCe*
G E N Y L * N N F V * L * E G D I P W Q L R I L * E I F * K D M D
aa cct ctc tta atg gag atc tta tta aaa caa att gaa att ctt cct cta tat acc gtc gac tct tat aac atc ctc tag aag atc ttt cta tac cc
10 20 30 40 50 60 70 80 90

c ggc cgt ctt cgt caa agc cct cca gac cgg cgc gcc cag ggg cgg gtg act cgc cgc cgt ttc ttc tca gcc ggc ccc cac atg gac gtc cgg cgg tcc < 200
G R L R Q S P P D R R A Q G P V T R R R F F S A G P H M D V R R S
g ccg gca gaa gca gtt tgg gga ggt ctg gcc gcg cgg gtc ccc gcc cac tga gcg cgg gca aag aag agt cgg ccg ggg gtg tac ctg cag gcc gcc agg
110 120 130 140 150 160 170 180 190

gcc aag cag acc ctt ctt cag tcc ttc aat agc atg aag cgt cgc agc gac ggc gtc caa cgt atc cac atg gag ggg ctg gca ttc aAA CGA GCG CTC C < 300
A K Q T L L Q S F N S M K R R S D G V Q R I H M E G L A F K R A L R
cgg ttc gtc tgg gaa gca gtt tgg gga ggt tta ttc gca gcg tgc ctg ccg cag gtt gca tag gtg tac ctc ccc gac cgt aag ttt GCT CGC GAG G
210 220 230 240 250 260 270 280 290

gg cct ccc agg tat tgc tgc gac gca tgc aga atc gcc cct tcc tcc aag tcc caa tcc tcc ctc tcc gac gcc aaa gag gac agc gag acc gat ttg ga < 400
P R R Y S S D A S R I A A S S K S Q S S L S D G K E D S F T D L D
cc gga gcg tcc ata agc agc ctc cgt agt gct tag cgg cga agg agg ttc agc gtt agg agc gag agc ctc cgg ttt ctc ctc tgc tgc cta aac ct
310 320 330 340 350 360 370 380 390

T GAC GAC TGC AGC GAT GAC TGC ATG ATC GAC GGA TAT CTG GCG GCC CCC TCA GGT CCG ATT CCG CCG ACG ACG TCC GAG GAT CCC GGT ATC GTG GTC ATC < 500
D D C S D A S M I D G Y L A A P S G R I R R T T S E D P G I V V I
A CTG CTG ACG TCG CTA GCG AGC TAC TAG CTG CCT ATA GAC CCG GGG AGT CCA GCC TAA GCC GCG TGC TGC AGG CTC CTA GGG CCA TAG CAC CAG TAG
410 420 430 440 450 460 470 480 490

GAG CCC TCT GGT CCA CAC ACG CAC ACC GCC ATT CTA TTG CAC GGC ATG TAC TGC TCT CCG GAG AGC TCC GAT ACA ATT GTC GGT CTA CCT GCT GCT GTC A < 600
E F S G F H T H T A I L L H G M Y C S E S S D T I V G L F A A V K
CTC GGG AGA CCA GGT GTC TGC GTG TGG CGG TAA GAT AAC GTC CCG TAC ATG ACG AGA GGC CTC TCG AGG CTA TGT TAA CAG CCA GAT GGA CGA CAG T
510 520 530 540 550 560 570 580 590

aa ttc ctt ggc ctg ctg cca ggc att aat tat gtc ttc cng cat gcg cgg cgt cng acc atc agc tgc cng aca ggc ccc gag gca aac gtt gcc ttc tg < 700
S L G L L P G I K Y V F P H A P R R T I S W P T G P E A N G A S W
TT AGG GAA CCG GAC GAC GGT CCC TAA TTT ATA CAG AAG GGC GTA CCG GGC GCA GCC TGG TAG TCG ACC GGC TGT CCG GGG CTC CGT TTG CAA CCG AGG AC
610 620 630 640 650 660 670 680 690

G TAC AAT TAC TAC ACG CCG CCG GAT GGC GAG CTC GAG CAT GAC GTC TTG AAT GAG GCG CAC CTG GCT TCG CAA ACC CCG CCG ATC CAC TCC ATT GTC GAG < 800
Y N Y Y T R R D G E L E H D V L N E A H L A S Q T R R I H S I V E
C ATG TTA ATG ATG TGC GCG GCC CTA CCG CTC GAG CTC GTA CTG CAG AAC TTA CTC CCG GTG GAC CGA AGC GTT TGG GCC TAG GTG AGG TAA CAG CTC
710 720 730 740 750 760 770 780 790

CGC GAG GCT GCG CTT CTT GGG GGC GAC GCG CCG CCG ATC ATG CTC GGT GGG AGC TCA CAA GGG GGC ACA GTC GCT CTC CAT GCT ACG ATC AAT TAC CAT C < 900
R E A A L L L G G D A R R I M L G G S S Q G G T V A L H A T I N Y H R
GCG CTC CGA CCG GAA GAA CCC CCG CTG CCG GCG GCG TAG TAC GAG CCA CCG TCG AGT GTT CCC CCG TGT CAG CGA GAG GTA CGA TGC TAG TTA ATG GTA G
810 820 830 840 850 860 870 880 890

gg cct ctc ggc gca ctg ctt tgt ctt cgc tca tgt ctc gat tgc atg acg ttt cgg aga gac aag cgc tcc ccc gcc gcc ggc aca ccc gtc ttc gt < 1000
P L G A L L C L R S C L I D S V T F P R D K R S P A A G T P V F V
cc gga gag cgc cot gac gaa aca gaa cgt acg aca gag tag tgc cga cca ccc tgc agt gtt ccc cgg tgt cag cga gag gta cga tgc tag tta atg gta g
910 920 930 940 950 960 970 980 990

c ttc gcc gcc gcc cag cag aag gtg tat gcc cca cag ctg cag tat cgc gcc ttc tct ctg ctg gcc gac tct ggc ttc cag gtg gag tgg cat gtg gaa < 1100
F A A A Q D K V Y A P Q L Q Y R G F S L L A D S G F H V E W H V E
g aag cgg cgg cgg gtc ttc cac ata cgg ggt gtc gac gtc ata cgg aag aga gac gac cgg ctg aga cgg aag gtg cag ctc acc gta cac ctt
1010 1020 1030 1040 1050 1060 1070 1080 1090

cct cat ctg acg cac tgg gac gag tgc cgc aca gag aaa cgc gta gtc gct gca tgg atc gcg cgc aca gtc cga gag cgc ggg ctc cga gag cgc gag c < 1200
P H L T H W D E S R T E K R V V A A W I A R T V R E R G L R E R E R
gga gta gac tgc acc ctg ctc agc cgg tgt ctc ttt gcg cat cag cga cgt acc tag cgc gcg tgt cag gct ctc gcg ccc gag gct ctc gcg ctc g
1110 1120 1130 1140 1150 1160 1170 1180 1190

NotI
|

gc tcc gta tgc cct gag ctt atc ctg gag cct ctt ttc atc ttg ctg aaa aac tgc agc cat ccg gaa gat ctg gcg gcc gca ctc gag cac cac cac ca < 1300
S V S P E L I L E P L F I L L K N S S H P E D L A A A L E H H H H
cg agg cat agc gga ctc gaa tag gac ctc gga gaa aag tag aac gac ttt ttg agc tgc gta ggc ctt cta gac cgc cgg cot gac gac ctc gta gta gt
1210 1220 1230 1240 1250 1260 1270 1280 1290

c cac cac tga gat ccg gct gct aac aaa gcc cga aag aat aca gt < 1345
H H * D P A A N K A R K N T X
g gtg gtg act cta ggc cga cga ttg ttt cgg gct ttc tta tgt ca
1310 1320 1330 1340

Fin d'*IgTeCe* dans *pET28-IgTeCe*

Annexe 12 : Séquençage du plasmide *pET52-IgTeCe*

NcoI
|

TC CCC TTC TAG AAT AAT TTT GTT TAA CTT TAA GAA GGA GAT ATA CCA TGG CAG CTG AGA ATA TTG TAG GAG ATC TTC TAG AAA GAT ATG GGC GGC CGT CT < 100 Début d'*IgTeCe*
 P F * N N F V * L * E G D I P W Q L R I L * E I F * K D M G G R L
 AG GGG AAG ATC TTA TTA AAA CAA ATT GAA ATT CTT CCT CTA TAT GGT ACC GTC GAC TCT TAT AAC ATC CTC TAG AAG ATC TTT CTA TAC CCG CCG GCA GA
 10 20 30 40 50 60 70 80 90

T CGT CAA AGC CCT CCA GAC CGG CGC GCC CAG GGG CCG GTG ACT CGC CGC CGT TTC TTC TCA GCC GGC CCC CAC ATG GAC GTC CGG CCG TCC GCC AAG CAG < 200
 R Q S P P D R R A Q G P V T R R R P F S A G P H M D V R R S A K Q
 A GCA GTT TCG GGA GGT CTG GCC GCG CCG GTC CCC GGC CAC TGA GCG GCG GCA AAG AAG AGT CGG CCG GGG GTG TAC CTG CAG GCC GCC AGG CCG TTC GTC
 110 120 130 140 150 160 170 180 190

ACC CTT CTT CAG TCC TTC AAT AGC ATG AAG CGT CGC AGC GAC GGC GTC CAA CGT ATC CAC ATG GAG GGG CTG GCA TTC AAA CGA GCG CTC CGG CCT CGC A < 300
 T L L Q S P N S M K R R S D G V Q R I H M E G L A F K R A L R P R R
 TGG GAA GAA GTC AAG AAG TTA TCG TAC TTC GCA GCG TCG CTG CCG CAG GTT GCA TAG GTG TAC CTC CCC GAC CGT AAG TTT GCT CGC GAG GCC GGA GCG T
 210 220 230 240 250 260 270 280 290

GG TAT TCG TCG GAC GCA TCC AGA ATC GCC GCT TCC TCC AAG TCG CAA TCC TCG CTC TCG GAC GGC AAA GAG GAC AGC GAG ACC GAT TTG GAT GAC GAC TG < 400
 G P H T H T A T L L H G M Y C S P E S S D T I V G L P A A V K S L G
 CC ATA AGC AGC CTG CGT AGG TCT TAG CCG CGA AGG AGG TTC AGC GTT AGG AGC GAG AGC CTG CCG TTT CTC CTG TCG CTC TGG CTA AAC CTA CTG CTG AC

C AGC GAT GCC TCG ATG ATC GAC GGA TAT CTG GCG GCC CCC TCA GGT CCG ATT CCG CCG ACG ACG TCC GAG GAT CCC GGT ATC GTG GTC ATC GAC CCC TCT < 500
 S D A S M I D G Y L A A P S G R I R R T T S E D P G I V V I E P S
 G TCG CTA CCG AGC TAC TAG CTG CCT ATA GAC CGC CCG GGG AGT CCA GCC TAA GCC CCG TGC TGC AGG CTC CTA GGG CCA TAG CAC CAG TAG CTC GGG AGA
 410 420 430 440 450 460 470 480 490

GGT CCA CAC ACG CAC ACC GGC ATT CTA TTG CAC GGC ATG TAC TGC TCT CCG GAG AGC TCC GAT ACA ATT GTC GGT CTA CCT GCT GCT GTC AAA TCC CTT G < 600
 G P H T H T A T L L H G M Y C S P E S S D T I V G L P A A V K S L G
 CCA GGT GTG TGC GTG TGG CCG TAA GAT AAC GTG CCG TAC ATG ACG AGA GGC CTC TCG AGG CTA TGT TAA CAG CCA GAT GGA CGA CGA CAG TTT AGG GAA C
 510 520 530 540 550 560 570 580 590

GC CTG CTG CCA GGG ATT AAA TAT GTC TTC CCG CAT GCG CCG CGT CCG ACC ATC AGC TGG CCG ACA GGC CCC GAG GCA AAC GTT GGC TCC TGG TAC AAT TA < 700
 L L P G I K Y V F P H A P R R R T I S W P T G P E A N V A S W Y N Y
 CG GAC GAC GGT CCC TAA TTT ATA CAG AAG GGC GTA CCG GGC GCA GCC TGG TAG TCG ACC GGC TGT CCG GGG CTC CGT TTG CAA CCG AGG ACC ATG TTA AT
 610 620 630 640 650 660 670 680 690

C TAC ACG CCG CCG GAT GGC GAG CTC GAG CAT GAC GTC TTG AAT GAG GCG CAC CTG GCT TCG CAA ACC CCG CCG ATC CAC TCC ATT GTC GAG CCG GCT < 800
 Y T R R R D G E L E H D V L N E A H L A S Q T R R R I H S I V E R E A
 G ATG TGC GCG GCC CTA CCG CTC GAG CTC GTA CTG CAG AAC TTA CTC CGC GTG GAC CGA AGC GTT TGG GCC GGC TAG GTG AGG TAA CAG CTC GCG CTC CGA
 710 720 730 740 750 760 770 780 790

GCG CTT CTT GGG GGC GAC GCG CCG CCG ATC ATG CTC GGT GGG AGC TCA CAA GGG GGC ACA GTC GCT CTC CAT GCT ACG ATC AAT TAC CAT CCG CCT CTC G < 900
 A L L G G D A R R I M L G G S S Q G G T V A L H A T I N Y H R P L G
 CCG GAA GAA CCC CCG CTG CCG GCG CCG TAG TAC GAG CCA CCC TCG AGT GTT CCC CCG TGT CAG CGA GAG GTA CGA TGC TAG TTA ATG GTA GCC GGA GAG C
 810 820 830 840 850 860 870 880 890

GC GCA CTG CTT TGT CTT CGC TCA TGT CTC ATC GAT TCG GTC ACG TTT CCG AGA GAC AAG CCG TCC CCC GGC GGC ACA CCC GTC TTC GTC TTC GCG GC < 1000
 A L L C L R S C L I D S V T F P R R D K R S P A A G T P V F V F A A
 CG CGT GAC GAA ACA GAA GCG AGT ACA GAG TAG CTA AGC CAG TGC AAA GGC TCT CTG TTC GCG AGG GGG CCG CCG CCG TGT GGG CAG AAG CAG AAG CCG CG
 910 920 930 940 950 960 970 980 990

```

C GCC CAG GAC AAG GTG TAT GCC CCA CAG CTG CAG TAT CGC GGC TTC TCT CTG CTG GCC GAC TCT GGC TTC CAC GTG GAG TGG CAT GTG GAA CCT CAT CTG < 1100
A Q D K V Y A P Q L Q Y R G F S L L A D S G F H V E W H V E P H L
G CGG GTC CTG TTC CAC ATA CGG GGT GTC GAC GTC ATA GCG CCG AAG AGA GAC GAC CGG CTG AGA CCG AAG GTG CAC CTC ACC GTA CAC CTT GGA GTA GAC
1010 1020 1030 1040 1050 1060 1070 1080 1090

```

```

ACG CAC TGG GAC GAG TCG CGC ACA GAG AAA CGC GTA GTC GCT GCA TGG ATC GCG CGC ACA GTC CGA GAG CGC GGG CTC CGA GAG CGC GAG CGC TCC GTA T < 1200
T H W D E S R T E K R V V A A W I A R T V R E R G L R E R E R S V S
TGC GTG ACC CTG CTC AGC GCG TGT CTC TTT GCG CAT CAG CGA CGT ACC TAG CGC GCG TGT CAG GCT CTC GCG CCC GAG GCT CTC GCG CTC GCG AGG CAT A
1110 1120 1130 1140 1150 1160 1170 1180 1190

```

```

CG CCT GAG CTT ATC CTG GAG COT CTT TTC ATC TTG CTG AAA AAC TCG AGC CAT CCG GAA GAT CTG GCG GCC GCA GAG CTC GCT CTG GTG CCA CGC GGT AG < 1300
P E L I L E F L F I L L K N S S H F E D L A A A E L A L V F R G S
GC GGA CTC GAA TAG GAC CTC GGA GAA AAG TAG AAC GAC TTT TTG AGC TCG GTA GGC CTT CTA GAC CGC CCG CGT CTC GAG CGA GAC CAC GGT GCG CCA TC
1210 1220 1230 1240 1250 1260 1270 1280 1290

```

```

>10H1s
T TCC GCT CAT CAC CAC CAT CAT CAC CAT CAC CAC CAG A < 1338
S A H H H H H H H H Q X
A AGG CGA GTA CTC GTA GTA GTA GTC GTC GTC T
1310 1320 1330

```

Fin d'IgTeCe dans pET52-IgTeCe

BIBLIOGRAPHIE

BIBLIOGRAPHIE

Thèses

- Bergé J.-P. (1996).** Optimisation de la production en masse de *Skeletonema costatum* (Grev.) Cleve. sur une eau souterraine salée en vue de sa valorisation: caractérisation des lipides et recherche de substances à activité biologique. Thèse de doctorat. Université de Nantes.
- Devos M. (2005).** Applications des lipases pour la valorisation de l'acide docosahexaénoïque de la microalgue *Isochrysis galbana*. Thèse de doctorat. Département Génie Biologique, IUT de Laval, Université du Maine.
- Godet S. (2008).** Isolement et caractérisation de trois gènes codant une lipase et deux estérases hypothétiques chez la microalgue marine *Isochrysis galbana* (Prymnesiophyceae, Haptophyta). Thèse de doctorat. Département Génie Biologique, IUT de Laval, Université du Maine.
- Poisson L. (1999).** Valorisation enzymatique de lipides: 1. Synthèse de cires à partir de la matière grasse laitière 2. Production et traitements d'acides gras polyinsaturés de microalgues. Thèse de doctorat. Département Génie Biologique, IUT de Laval, Université du Maine.

Publications

- Addou, S., Rentzsch, R., Lee, D., and Orengo, C.A. (2009).** Domain-based and family-specific sequence identity thresholds increase the levels of reliable protein function transfer. *Journal of molecular biology* 387, 416-430.
- Akbari, N., Khajeh, K., Rezaie, S., Mirdamadi, S., Shavandi, M., and Ghaemi, N. (2010).** High-level expression of lipase in *Escherichia coli* and recovery of active recombinant enzyme through in vitro refolding. *Protein expression and purification* 70, 75-80.
- Akoh, C.C., Lee, G.C., Liaw, Y.C., Huang, T.H., and Shaw, J.F. (2004).** GDSL family of serine esterases/lipases. *Progress in lipid research* 43, 534-552.
- Ali, Y.B., Verger, R., and Abousalham, A. (2012).** Lipases or esterases: does it really matter? Toward a new bio-physico-chemical classification. *Methods in molecular biology* 861, 31-51.
- Aloulou, A., Ali, Y.B., Bezzine, S., Gargouri, Y., and Gelb, M.H. (2012).** Phospholipases: an overview. *Methods in molecular biology* 861, 63-85.
- Aloulou, A., Rodriguez, J.A., Fernandez, S., van Oosterhout, D., Puccinelli, D., and Carriere, F. (2006).** Exploring the specific features of interfacial enzymology based on lipase studies. *Biochimica et biophysica acta* 1761, 995-1013.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990).** Basic local alignment search tool. *Journal of molecular biology* 215, 403-410.
- Alvarez-Macarie, E., Augier-Magro, V., and Baratti, J. (1999).** Characterization of a thermostable esterase activity from the moderate thermophile *Bacillus licheniformis*. *Bioscience, biotechnology, and biochemistry* 63, 1865-1870.
- Andersen, C.L., Matthey-Dupraz, A., Missiakas, D., and Raina, S. (1997).** A new *Escherichia coli* gene, dsbG, encodes a periplasmic protein involved in disulphide bond formation, required for recycling DsbA/DsbB and DsbC redox proteins. *Molecular Microbiology* 26, 121-132.
- Anderson, E.M., Larsson, K.M., and Kirk, O. (1998).** One Biocatalyst—Many Applications: The Use of *Candida Antarctica* B-Lipase in Organic Synthesis. *Biocatalysis and Biotransformation* 16, 181-204.
- Angov, E. (2011).** Codon usage: nature's roadmap to expression and folding of proteins. *Biotechnology journal* 6, 650-659.
- Antia, N.J., Bilinski, E., and Lau, Y.C. (1970).** Identification and characterization of phospholipase D in a unicellular red alga (*Porphyridium cruentum*). *Canadian journal of biochemistry* 48, 643-648.
- Aoubala, M., Ivanova, M., Douchet, I., De Caro, A., and Verger, R. (1995).** Interfacial binding of human gastric lipase to lipid monolayers, measured with an ELISA. *Biochemistry* 34, 10786-10793.

- Arisz, S.A., and Munnik, T. (2011).** The salt stress-induced LPA response in *Chlamydomonas* is produced via PLA(2) hydrolysis of DGK-generated phosphatidic acid. *Journal of lipid research* 52, 2012-2020.
- Arnau, J., Lauritzen, C., Petersen, G.E., and Pedersen, J. (2006).** Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins. *Protein expression and purification* 48, 1-13.
- Arpigny, J.L., Feller, G., and Gerday, C. (1995).** Corrigendum to "Cloning, sequence and structural features of a lipase from the antarctic facultative psychrophile *Psychrobacter immobilis* B10" [Biochim. Biophys. Acta 1171 (1993) 331-333]. *Biochimica et biophysica acta* 1263, 103.
- Arpigny, J.L., and Jaeger, K.E. (1999).** Bacterial lipolytic enzymes: classification and properties. *The Biochemical journal* 343 Pt 1, 177-183.
- Arrese, E.L., Patel, R.T., and Soulages, J.L. (2006).** The main triglyceride-lipase from the insect fat body is an active phospholipase A(1): identification and characterization. *Journal of lipid research* 47, 2656-2667.
- Arrese, E.L., and Soulages, J.L. (2010).** Insect fat body: energy, metabolism, and regulation. *Annual review of entomology* 55, 207-225.
- Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., et al. (2012).** ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Research* 40, W597-W603.
- Azaman, S.N., Ramakrishnan, N.R., Tan, J.S., Rahim, R.A., Abdullah, M.P., and Ariff, A.B. (2010).** Optimization of an induction strategy for improving interferon-alpha2b production in the periplasm of *Escherichia coli* using response surface methodology. *Biotechnology and applied biochemistry* 56, 141-150.
- Bairoch, A., and Apweiler, R. (2000).** The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28, 45-48.
- Balashev, K., Jensen, T.R., Kjaer, K., and Bjornholm, T. (2001).** Novel methods for studying lipids and lipases and their mutual interaction at interfaces. Part I. Atomic force microscopy. *Biochimie* 83, 387-397.
- Bamji-Mirza, M., and Yao, Z. (2011).** PHOSPHOLIPASES. In *Animal Lipid Biochemistry*.
- Becher, D., Büttner, K., Moche, M., Heßling, B., and Hecker, M. (2011).** From the genome sequence to the protein inventory of *Bacillus subtilis*. *Proteomics* 11, 2971-2980.
- Beer, H.D., Wohlfahrt, G., McCarthy, J.E., Schomburg, D., and Schmid, R.D. (1996).** Analysis of the catalytic mechanism of a fungal lipase using computer-aided design and structural mutants. *Protein engineering* 9, 507-517.
- Beisson, F., Ferte, N., Nari, J., Noat, G., Arondel, V., and Verger, R. (1999).** Use of naturally fluorescent triacylglycerols from *Parinari glaberrimum* to detect low lipase activities from *Arabidopsis thaliana* seedlings. *Journal of lipid research* 40, 2313-2321.
- Beisson, F., Tiss, A., Rivière, C., and Verger, R. (2000).** Methods for lipase detection and assay: a critical review. *European Journal of Lipid Science and Technology* 102, 133-153.
- Benzonana, G., and Desnuelle, P. (1965).** [Kinetic study of the action of pancreatic lipase on emulsified triglycerides. Enzymology assay in heterogeneous medium]. *Biochimica et biophysica acta* 105, 121-136.
- Bereuter, T.L., and Lorbeer, E. (1995).** Monitoring of lipase-catalyzed cleavage of acylglycerols by high-temperature gas chromatography. *Journal of Chromatography A* 697, 469-474.
- Berrow, N.S., Bussow, K., Coutard, B., Diprose, J., Ekberg, M., Folkers, G.E., Levy, N., Lieu, V., Owens, R.J., Peleg, Y., et al. (2006).** Recombinant protein expression and solubility screening in *Escherichia coli*: a comparative study. *Acta crystallographica Section D, Biological crystallography* 62, 1218-1226.
- Bessette, P.H., Aslund, F., Beckwith, J., and Georgiou, G. (1999).** Efficient folding of proteins with multiple disulfide bonds in the *Escherichia coli* cytoplasm. *Proceedings of the National Academy of Sciences of the United States of America* 96, 13703-13708.
- Bilinski, E., Antia, N.J., and Lau, Y.C. (1968).** Characterization of phospholipase C from a marine planktonic alga (*Monochrysis lutheri*). *Biochimica et biophysica acta* 159, 496-502.
- Birner-Gruenberger, R., Susani-Etzerodt, H., Waldhuber, M., Riesenhuber, G., Schmidinger, H., Rechberger, G., Kollroser, M., Strauss, J.G., Lass, A., Zimmermann, R., et al. (2005).** The Lipolytic Proteome of Mouse Adipose Tissue. *Molecular & Cellular Proteomics* 4, 1710-1717.

- Blank, K., Morfill, J., Gumpp, H., and Gaub, H.E. (2006). Functional expression of *Candida antarctica* lipase B in *Escherichia coli*. *Journal of biotechnology* 125, 474-483.
- Boer, E., Steinborn, G., Kunze, G., and Gellissen, G. (2007). Yeast expression platforms. *Applied microbiology and biotechnology* 77, 513-523.
- Boratyn, G.M., Schaffer, A.A., Agarwala, R., Altschul, S.F., Lipman, D.J., and Madden, T.L. (2012). Domain enhanced lookup time accelerated BLAST. *Biology direct* 7, 12.
- Borgström, B., and Brockman, H.L. (1984). *Lipases* (Elsevier).
- Borgstrom, B., and Erlanson, C. (1973). Pancreatic lipase and co-lipase. Interactions and effects of bile salts and other detergents. *European journal of biochemistry / FEBS* 37, 60-68.
- Bornscheuer, U.T. (1995). Lipase-catalyzed syntheses of monoacylglycerols. *Enzyme and Microbial Technology* 17, 578-586.
- Bornscheuer, U.T., and Kazlauskas, R.J. (1999). *Hydrolases in organic synthesis* (Wiley-VCH Weinheim).
- Bottcher, D., Brusehaber, E., Doderer, K., and Bornscheuer, U.T. (2007). Functional expression of the gamma-isoenzyme of pig liver carboxyl esterase in *Escherichia coli*. *Applied microbiology and biotechnology* 73, 1282-1289.
- Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* 72, 248-254.
- Brake, A.J., Merryweather, J.P., Coit, D.G., Heberlein, U.A., Masiarz, F.R., Mullenbach, G.T., Urdea, M.S., Valenzuela, P., and Barr, P.J. (1984). Alpha-factor-directed synthesis and secretion of mature foreign proteins in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences* 81, 4642-4646.
- Brehelin, C., Kessler, F., and van Wijk, K.J. (2007). Plastoglobules: versatile lipoprotein particles in plastids. *Trends in plant science* 12, 260-266.
- Briquet-Laugier, V., Ben-Zeev, O., and Doolittle, M.H. (1999). Determining lipoprotein lipase and hepatic lipase activity using radiolabeled substrates. *Methods in molecular biology* 109, 81-94.
- Brocker, C., Carpenter, C., Nebert, D.W., and Vasiliou, V. (2010). Evolutionary divergence and functions of the human acyl-CoA thioesterase gene (ACOT) family. *Human genomics* 4, 411-420.
- Brondyk, W.H. (2009). Selecting an appropriate method for expressing a recombinant protein. *Methods in enzymology* 463, 131-147.
- Brzozowski, A.M., Derewenda, U., Derewenda, Z.S., Dodson, G.G., Lawson, D.M., Turkenburg, J.P., Bjorkling, F., Huge-Jensen, B., Patkar, S.A., and Thim, L. (1991). A model for interfacial activation in lipases from the structure of a fungal lipase-inhibitor complex. *Nature* 351, 491-494.
- Burgess-Brown, N.A., Sharma, S., Sobott, F., Loenarz, C., Oppermann, U., and Gileadi, O. (2008). Codon optimization can improve expression of human genes in *Escherichia coli*: A multi-gene study. *Protein expression and purification* 59, 94-102.
- Cadore, J.P., and Bernard, O. (2008). [Lipid biofuel production with microalgae: potential and challenges]. *Journal de la Societe de biologie* 202, 201-211.
- Canaan, S., Roussel, A., Verger, R., and Cambillau, C. (1999). Gastric lipase: crystal structure and activity. *Biochimica et biophysica acta* 1441, 197-204.
- Canonne, J., Froidure-Nicolas, S., and Rivas, S. (2011). Phospholipases in action during plant defense signaling. *Plant signaling & behavior* 6, 13-18.
- Cantu, D.C., Chen, Y., and Reilly, P.J. (2010). Thioesterases: a new perspective based on their primary and tertiary structures. *Protein science : a publication of the Protein Society* 19, 1281-1295.
- Casas-Godoy, L., Duquesne, S., Bordes, F., Sandoval, G., and Marty, A. (2012). Lipases: An Overview. In *Lipases and Phospholipases*, G. Sandoval, ed. (Humana Press), pp. 3-30.
- Celik, E., and Calik, P. (2012). Production of recombinant proteins by yeast cells. *Biotechnology advances* 30, 1108-1118.
- Ceroni, A., Passerini, A., Vullo, A., and Frasconi, P. (2006). DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. *Nucleic Acids Res* 34, W177-181.
- Chahinian, H., Ali, Y.B., Abousalham, A., Petry, S., Mandrich, L., Manco, G., Canaan, S., and Sarda, L. (2005). Substrate specificity and kinetic properties of enzymes belonging to the hormone-sensitive lipase

- family: comparison with non-lipolytic and lipolytic carboxylesterases. *Biochimica et biophysica acta* 1738, 29-36.
- Chan, Y.A., Podevels, A.M., Kevany, B.M., and Thomas, M.G. (2009).** Biosynthesis of polyketide synthase extender units. *Natural product reports* 26, 90-114.
- Chang, R.-C., Chou, S.-J., and Shaw, J.-F. (1994).** Multiple forms and functions of *Candida rugosa* lipase. *Biotechnology and applied biochemistry* 19, 93-97.
- Chen, G., Peng, Z.Y., Shan, L., Xuan, N., Tang, G.Y., Zhang, Y., Li, L., He, Q.F., and Bi, Y.P. (2012).** Cloning of acyl-ACP thioesterase FatA from *Arachis hypogaea* L. and its expression in *Escherichia coli*. *Journal of biomedicine & biotechnology* 2012, 652579.
- Chen, J., Ishii, T., Shimura, S., Kirimura, K., and Usami, S. (1992).** Lipase production by *Trichosporon fermentans* WU-C12, a newly isolated yeast. *Journal of Fermentation and Bioengineering* 73, 412-414.
- Choi, Y.J., Miguez, C.B., and Lee, B.H. (2004).** Characterization and heterologous gene expression of a novel esterase from *Lactobacillus casei* CL96. *Applied and environmental microbiology* 70, 3213-3221.
- Chrisope, G., Fox, C., and Marshall, R. (1976).** Lecithin agar for detection of microbial phospholipases. *Applied and environmental microbiology* 31, 784-786.
- Christie, W.W., and Han, X. (2010).** Lipid Analysis - Isolation, Separation, Identification and Lipidomic Analysis, Vol 24, 4th edn (Oily Press, Bridgwater, U.K).
- Cihangir, N., and Sarikaya, E. (2004).** Investigation of lipase production by a new isolate of *Aspergillus* sp. *World Journal of Microbiology and Biotechnology* 20, 193-197.
- Cole, C., Barber, J.D., and Barton, G.J. (2008).** The Jpred 3 secondary structure prediction server. *Nucleic Acids Research* 36, W197-W201.
- Cousin, X., Hotelier, T., Lievin, P., Toutant, J.P., and Chatonnet, A. (1996).** A cholinesterase genes server (ESTHER): a database of cholinesterase-related sequences for multiple alignments, phylogenetic relationships, mutations and structural data retrieval. *Nucleic Acids Res* 24, 132-136.
- Cregg, J.M., Tolstorukov, I., Kusari, A., Sunga, J., Madden, K., and Chappell, T. (2009).** Expression in the yeast *Pichia pastoris*. *Methods in enzymology* 463, 169-189.
- Das, K.M., Banerjee, S., Shekhar, N., Damodaran, K., Nair, R., Somani, S., Raiker, V.P., Jain, S., and Padmanabhan, S. (2011).** Cloning, Soluble Expression and Purification of High Yield Recombinant hGMCSF in *Escherichia coli*. *International journal of molecular sciences* 12, 2064-2076.
- de Azeredo, L.A., Gomes, P.M., Sant'Anna, G.L., Jr., Castilho, L.R., and Freire, D.M. (2007).** Production and regulation of lipase activity from *Penicillium restrictum* in submerged and solid-state fermentations. *Current microbiology* 54, 361-365.
- De Caro, A., Bezzine, S., Lopez, V., Aoubala, M., Daniel, C., Verger, R., and Carriere, F. (1999).** Immunological characterization of digestive lipases. *Methods in molecular biology* 109, 239-256.
- Demain, A.L., and Vaishnav, P. (2009).** Production of recombinant proteins by microbes and higher organisms. *Biotechnology advances* 27, 297-306.
- Derewenda, U., Swenson, L., Wei, Y., Green, R., Kobos, P.M., Joerger, R., Haas, M.J., and Derewenda, Z.S. (1994).** Conformational lability of lipases observed in the absence of an oil-water interface: crystallographic studies of enzymes from the fungi *Humicola lanuginosa* and *Rhizopus delemar*. *Journal of lipid research* 35, 524-534.
- Derewenda, Z.S. (1994).** Structure and function of lipases. *Advances in protein chemistry* 45, 1-52.
- Derewenda, Z.S., Derewenda, U., and Dodson, G.G. (1992).** The crystal and molecular structure of the *Rhizomucor miehei* triacylglyceride lipase at 1.9 Å resolution. *Journal of molecular biology* 227, 818-839.
- Dillon, S.C., and Bateman, A. (2004).** The Hotdog fold: wrapping up a superfamily of thioesterases and dehydratases. *BMC bioinformatics* 5, 109.
- Docherty, A.J.P., Bodmer, M.W., Angal, S., Verger, R., Riviere, C., Lowe, P.A., Lyons, A., Emtage, J.S., and Harris, T.J.R. (1985).** Molecular cloning and nucleotide sequence of rat lingual lipase cDNA. *Nucleic Acids Research* 13, 1891-1903.
- Dolan, J.W., Bell, A.C., Hube, B., Schaller, M., Warner, T.F., and Balish, E. (2004).** *Candida albicans* PLD I activity is required for full virulence. *Medical mycology : official publication of the International Society for Human and Animal Mycology* 42, 439-447.

- Dujon, B., Sherman, D., Fischer, G., Durrens, P., Casaregola, S., Lafontaine, I., De Montigny, J., Marck, C., Neuveglise, C., Talla, E., *et al.* (2004). Genome evolution in yeasts. *Nature* 430, 35-44.
- Duncan, J.A., and Gilman, A.G. (1998). A cytoplasmic acyl-protein thioesterase that removes palmitate from G protein alpha subunits and p21(RAS). *The Journal of biological chemistry* 273, 15830-15837.
- Durocher, Y., Perret, S., and Kamen, A. (2002). High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells. *Nucleic Acids Res* 30, E9.
- Dutra, J.C., da, C.T.S., Bevilaqua, J.V., Damaso, M.C., Couri, S., Langone, M.A., and Senna, L.F. (2008). Lipase production in solid-state fermentation monitoring biomass growth of *Aspergillus niger* using digital image processing. *Applied biochemistry and biotechnology* 147, 63-75.
- Einspahr, K.J., Peeler, T.C., and Thompson, G.A. (1989). Phosphatidylinositol 4,5-Bisphosphate Phospholipase C and Phosphomonoesterase in *Dunaliella salina* Membranes. *Plant physiology* 90, 1115-1120.
- Ellaiah, P., Prabhakar, T., Ramakrishna, B., Taleb, A.T., and Adinarayana, K. (2004). Production of lipase by immobilized cells of *Aspergillus niger*. *Process Biochemistry* 39, 525-528.
- Emanuelsson, O., Nielsen, H., and von Heijne, G. (1999). ChloroP, a neural network-based method for predicting chloroplast transit peptides and their cleavage sites. *Protein science : a publication of the Protein Society* 8, 978-984.
- Engel, K.-H. (1992). Lipase-catalyzed enantioselective acidolysis of chiral 2-methylalkanoates. *J Am Oil Chem Soc* 69, 146-150.
- Ergan, F., and Andre, G. (1989). Simple high performance liquid chromatography methods for monitoring lipase reactions. *Lipids* 24, 76-78.
- Esposito, D., and Chatterjee, D.K. (2006). Enhancement of soluble protein expression through the use of fusion tags. *Current opinion in biotechnology* 17, 353-358.
- Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.-y., Pieper, U., and Sali, A. (2001). Comparative Protein Structure Modeling Using MODELLER. In *Current Protocols in Protein Science* (John Wiley & Sons, Inc.).
- Fickers, P., Destain, J., and Thonart, P. (2008). Les lipases sont des hydrolases atypiques : principales caractéristiques et applications, Vol 12 (Gembloux, BELGIQUE: Facult; universitaire des sciences agronomiques).
- Fischer, M. (2003). The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Research* 31, 319-321.
- Franz, A.K., Danielewicz, M.A., Wong, D.M., Anderson, L.A., and Boothe, J.R. (2013). Phenotypic Screening with Oleaginous Microalgae Reveals Modulators of Lipid Productivity. *ACS Chemical Biology* 8, 1053-1062.
- Freydank, A.-C., Brandt, W., and Dräger, B. (2008). Protein structure modeling indicates hexahistidine-tag interference with enzyme activity. *Proteins: Structure, Function, and Bioinformatics* 72, 173-183.
- Fu, M. (2010). Codon usage bias in herpesvirus. *Archives of virology* 155, 391-396.
- Gade, G., and Auerswald, L. (2003). Mode of action of neuropeptides from the adipokinetic hormone family. *General and comparative endocrinology* 132, 10-20.
- Gaillardin, C. (2010). Lipases as Pathogenicity Factors of Fungi. In *Handbook of Hydrocarbon and Lipid Microbiology*, K. Timmis, ed. (Springer Berlin Heidelberg), pp. 3259-3268.
- Gandhi, N. (1997). Applications of lipase. *J Amer Oil Chem Soc* 74, 621-634.
- García-Fruitós, E., Vázquez, E., Díez-Gil, C., Corchero, J.L., Seras-Franzoso, J., Ratera, I., Veciana, J., and Villaverde, A. (2012). Bacterial inclusion bodies: making gold from waste. *Trends in biotechnology* 30, 65-70.
- Gargouri, Y., Pieroni, G., Riviere, C., Saunier, J.F., Lowe, P.A., Sarda, L., and Verger, R. (1986). Kinetic assay of human gastric lipase on short- and long-chain triacylglycerol emulsions. *Gastroenterology* 91, 919-925.

- Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodriguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., *et al.* (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microbial cell factories* 7, 11.
- Gavrilescu, M., and Chisti, Y. (2005). Biotechnology—a sustainable alternative for chemical industry. *Biotechnology advances* 23, 471-499.
- Geisse, S., and Fux, C. (2009). Recombinant protein production by transient gene transfer into Mammalian cells. *Methods in enzymology* 463, 223-238.
- Geszvain, K., and Tebo, B.M. (2010). Identification of a two-component regulatory pathway essential for Mn(II) oxidation in *Pseudomonas putida* GB-1. *Applied and environmental microbiology* 76, 1224-1231.
- Gil, F., Catalao, M.J., Moniz-Pereira, J., Leandro, P., McNeil, M., and Pimentel, M. (2008). The lytic cassette of mycobacteriophage Ms6 encodes an enzyme with lipolytic activity. *Microbiology* 154, 1364-1371.
- Gilbert, E.J. (1993). *Pseudomonas* lipases: biochemical properties and molecular cloning. *Enzyme Microb Technol* 15, 634-645.
- Gilbert, J.A., O'Dor, R., King, N., and Vogel, T.M. (2011). The importance of metagenomic surveys to microbial ecology: or why Darwin would have been a metagenomic scientist. *Microbial informatics and experimentation* 1, 5.
- Gilham, D., and Lehner, R. (2005). Techniques to measure lipase and esterase activity in vitro. *Methods* 36, 139-147.
- Gill, R.T., Valdes, J.J., and Bentley, W.E. (2000). A comparative study of global stress gene regulation in response to overexpression of recombinant proteins in *Escherichia coli*. *Metabolic engineering* 2, 178-189.
- Glogauer, A., Martini, V.P., Faoro, H., Couto, G.H., Muller-Santos, M., Monteiro, R.A., Mitchell, D.A., de Souza, E.M., Pedrosa, F.O., and Krieger, N. (2011). Identification and characterization of a new true lipase isolated through metagenomic approach. *Microbial cell factories* 10, 54.
- Godet, S., Hérault, J., Pencreac'h, G., Ergon, F., and Loiseau, C. (2012). Isolation and analysis of a gene from the marine microalga *Isochrysis galbana* that encodes a lipase-like protein. *Journal of Applied Phycology* 24, 1547-1553.
- Godet, S., Loiseau, C., Pencreac'h, G., Ergon, F., and Hérault, J. (2010). Isolation and sequence analysis of a cDNA encoding a novel putative esterase from the marine microalga *Isochrysis galbana* (*Prymnesiophyceae*, *Haptophyta*). *Journal of Phycology* 46, 679-684.
- Gong, Y., Guo, X., Wan, X., Liang, Z., and Jiang, M. (2011). Characterization of a novel thioesterase (PtTE) from *Phaeodactylum tricornutum*. *Journal of basic microbiology* 51, 666-672.
- Götz, F., Popp, F., Kom, E., and Schleifer, K.H. (1985). Complete nucleotide sequence of the lipase gene from *Staphylococcus hyicus* cloned in *Staphylococcus carnosus*. *Nucleic Acids Research* 13, 5895-5906.
- Green, M.R., Sambrook, J., and Sambrook, J. (2012). *Molecular cloning : a laboratory manual*, 4th edn (Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press).
- Griebeler, N., Polloni, A., Remonatto, D., Arbter, F., Vardanega, R., Cechet, J., Luccio, M., Oliveira, D., Treichel, H., Cansian, R., *et al.* (2011). Isolation and Screening of Lipase-Producing Fungi with Hydrolytic Activity. *Food and Bioprocess Technology* 4, 578-586.
- Griesbeck, C., Kobl, I., and Heitzer, M. (2006). *Chlamydomonas reinhardtii*. *Molecular biotechnology* 34, 213-223.
- Guiheneuf, F., Leu, S., Zarka, A., Khozin-Goldberg, I., Khalilov, I., and Boussiba, S. (2011). Cloning and molecular characterization of a novel acyl-CoA:diacylglycerol acyltransferase 1-like gene (PtDGAT1) from the diatom *Phaeodactylum tricornutum*. *The FEBS journal* 278, 3651-3666.
- Han, M.J., Lee, S.Y., and Hong, S.H. (2012). Comparative analysis of envelope proteomes in *Escherichia coli* B and K-12 strains. *J Microbiol Biotechnol* 22, 470-478.
- Harrison, R.L., and Jarvis, D.L. (2006). Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. *Advances in virus research* 68, 159-191.
- Hasan, F., Shah, A.A., and Hameed, A. (2006). Industrial applications of microbial lipases. *Enzyme and Microbial Technology* 39, 235-251.

- Hasan, F., Shah, A.A., and Hameed, A. (2009). Methods for detection and characterization of lipases: A comprehensive review. *Biotechnology advances* 27, 782-798.
- Hausmann, S., and Jaeger, K.E. (2010). Lipolytic Enzymes from Bacteria. In *Handbook of Hydrocarbon and Lipid Microbiology*, K. Timmis, ed. (Springer Berlin Heidelberg), pp. 1099-1126.
- Helmy, M., Lombard, S., and Pieroni, G. (1999). Ricin RCA60: evidence of its phospholipase activity. *Biochemical and biophysical research communications* 258, 252-255.
- Hermann, M., Kietzmann, M.U., Ivancic, M., Zenzmaier, C., Luiten, R.G., Skranc, W., Wubbolts, M., Winkler, M., Birner-Gruenberger, R., Pichler, H., *et al.* (2008). Alternative pig liver esterase (APLE) - cloning, identification and functional expression in *Pichia pastoris* of a versatile new biocatalyst. *Journal of biotechnology* 133, 301-310.
- Heyland, J., Fu, J., Blank, L.M., and Schmid, A. (2010). Quantitative physiology of *Pichia pastoris* during glucose-limited high-cell density fed-batch cultivation for recombinant protein production. *Biotechnology and bioengineering* 107, 357-368.
- Hills, M.J., and Mukherjee, K.D. (1988). Assay for triacylglycerol lipase by a rapid thin-layer chromatographic technique. *Journal of lipid research* 29, 1397-1399.
- Hodgson, A.L., Bird, P., and Nisbet, I.T. (1990). Cloning, nucleotide sequence, and expression in *Escherichia coli* of the phospholipase D gene from *Corynebacterium pseudotuberculosis*. *Journal of bacteriology* 172, 1256-1261.
- Hoffman, C.S., and Winston, F. (1987). A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. *Gene* 57, 267-272.
- Hofmann, A.F., and Borgstrom, B. (1963). Hydrolysis of long-chain monoglycerides in micellar solution by pancreatic lipase. *Biochimica et biophysica acta* 70, 317-331.
- Holmquist, M. (2000). Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms. *Current protein & peptide science* 1, 209-235.
- Horchani, H., Ben Salem, N., Chaari, A., Sayari, A., Gargouri, Y., and Verger, R. (2010). Staphylococcal lipases stereoselectively hydrolyse the sn-2 position of monomolecular films of diglyceride analogs. Application to sn-2 hydrolysis of triolein. *Journal of colloid and interface science* 347, 301-308.
- Ibrahim, A.S., Mirbod, F., Filler, S.G., Banno, Y., Cole, G.T., Kitajima, Y., Edwards, J.E., Nozawa, Y., and Ghannoum, M.A. (1995). Evidence implicating phospholipase as a virulence factor of *Candida albicans*. *Infection and immunity* 63, 1993-1998.
- Iijima, N., Tanaka, S., and Ota, Y. (1998). Purification and characterization of bile salt-activated lipase from the hepatopancreas of red sea bream, *Pagrus major*. *Fish physiology and biochemistry* 18, 59-69.
- Ileperuma, N.R., Marshall, S.D., Squire, C.J., Baker, H.M., Oakeshott, J.G., Russell, R.J., Plummer, K.M., Newcomb, R.D., and Baker, E.N. (2007). High-resolution crystal structure of plant carboxylesterase AeCXE1, from *Actinidia eriantha*, and its complex with a high-affinity inhibitor paraoxon. *Biochemistry* 46, 1851-1859.
- Irimescu, R.C.O.N.S.K.L.K.-M.H.-s.T., Furihata, K.C.O.N.S.K.L.K.-M.H.-s.T., Hata, K.C.O.N.S.K.L.K.-M.H.-s.T., and Yamane, T.W.C.C.-K.N.-S.A. (2010). Process for the production of glycerides with lipases (EP: EP-A- 0 882 797; JP-A- 1 019 042; JP-A- 62 228 290).
- Istivan, T.S., and Coloe, P.J. (2006). Phospholipase A in Gram-negative bacteria and its role in pathogenesis. *Microbiology* 152, 1263-1274.
- Itakura, K., Hirose, T., Crea, R., Riggs, A.D., Heyneker, H.L., Bolivar, F., and Boyer, H.W. (1977). Expression in *Escherichia coli* of a chemically synthesized gene for the hormone somatostatin. *Science* 198, 1056-1063.
- Ito, S., Kobayashi, T., Ara, K., Ozaki, K., Kawai, S., and Hatada, Y. (1998). Alkaline detergent enzymes from alkaliphiles: enzymatic properties, genetics, and structures. *Extremophiles : life under extreme conditions* 2, 185-190.
- Jacobs, A.C., Hood, I., Boyd, K.L., Olson, P.D., Morrison, J.M., Carson, S., Sayood, K., Iwen, P.C., Skaar, E.P., and Dunman, P.M. (2010). Inactivation of phospholipase D diminishes *Acinetobacter baumannii* pathogenesis. *Infection and immunity* 78, 1952-1962.

- Jaeger, K.-E., Ransac, S., Dijkstra, B.W., Colson, C., van Heuvel, M., and Misset, O. (1994). Bacterial lipases. *FEMS microbiology reviews* 15, 29-63.
- Jaeger, K.E., and Eggert, T. (2002). Lipases for biotechnology. *Current opinion in biotechnology* 13, 390-397.
- Jaeger, K.E., and Reetz, M.T. (1998). Microbial lipases form versatile tools for biotechnology. *Trends in biotechnology* 16, 396-403.
- Jaenicke, R. (1965). [Intermolecular Interaction in Heat Aggregation of Globular Proteins]. *Zeitschrift fur Naturforschung Teil B: Chemie, Biochemie, Biophysik, Biologie* 20, 21-31.
- Jenkins, N., Parekh, R.B., and James, D.C. (1996). Getting the glycosylation right: implications for the biotechnology industry. *Nature biotechnology* 14, 975-981.
- Jensen, R.G., deJong, F.A., and Clark, R.M. (1983). Determination of lipase specificity. *Lipids* 18, 239-252.
- Jette, J.F., and Ziomek, E. (1994). Determination of lipase activity by a rhodamine-triglyceride-agarose assay. *Analytical biochemistry* 219, 256-260.
- Jha, S.S., Jha, J.K., Chattopadhyaya, B., Basu, A., Sen, S.K., and Maiti, M.K. (2010). Cloning and characterization of cDNAs encoding for long-chain saturated acyl-ACP thioesterases from the developing seeds of *Brassica juncea*. *Plant physiology and biochemistry : PPB / Societe francaise de physiologie vegetale* 48, 476-480.
- Jiang, F., and Mannervik, B. (1999). Optimized heterologous expression of glutathione reductase from *Cyanobacterium anabaena* PCC 7120 and characterization of the recombinant protein. *Protein expression and purification* 15, 92-98.
- Johnston, M., and Davis, R.W. (1984). Sequences that regulate the divergent GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Molecular and cellular biology* 4, 1440-1448.
- Jones, J.M., Nau, K., Geraghty, M.T., Erdmann, R., and Gould, S.J. (1999). Identification of peroxisomal acyl-CoA thioesterases in yeast and humans. *The Journal of biological chemistry* 274, 9216-9223.
- Kang, H.Y., Kim, J.F., Kim, M.H., Park, S.H., Oh, T.K., and Hur, C.G. (2006). MELDB: a database for microbial esterases and lipases. *FEBS letters* 580, 2736-2740.
- Keeling, P.J. (2006).
- Kohler, G.A., Brenot, A., Haas-Stapleton, E., Agabian, N., Deva, R., and Nigam, S. (2006). Phospholipase A2 and phospholipase B activities in fungi. *Biochimica et biophysica acta* 1761, 1391-1399.
- Korbsrisate, S., Tomaras, A.P., Damnin, S., Ckumdee, J., Srinon, V., Lengwehasatit, I., Vasil, M.L., and Suparak, S. (2007). Characterization of two distinct phospholipase C enzymes from *Burkholderia pseudomallei*. *Microbiology* 153, 1907-1915.
- Koshland, D.E. (1958). Application of a Theory of Enzyme Specificity to Protein Synthesis. *Proceedings of the National Academy of Sciences of the United States of America* 44, 98-104.
- Kouker, G., and Jaeger, K.E. (1987). Specific and sensitive plate assay for bacterial lipases. *Applied and environmental microbiology* 53, 211-213.
- Kourist, R., Jochens, H., Bartsch, S., Kuipers, R., Padhi, S.K., Gall, M., Böttcher, D., Joosten, H.-J., and Bornscheuer, U.T. (2010). The α/β -Hydrolase Fold 3DM Database (ABHDB) as a Tool for Protein Engineering. *Chembiochem : a European journal of chemical biology* 11, 1635-1643.
- Krügenger, S., Zelena, K., Zorn, H., Nimtz, M., and Berger, R.G. (2009). Heterologous expression of an extra-cellular lipase from the basidiomycete *Pleurotus sapidus*. *Journal of Molecular Catalysis B: Enzymatic* 57, 16-21.
- Kück, U., and Hoff, B. (2010). New tools for the genetic manipulation of filamentous fungi. *Applied microbiology and biotechnology* 86, 51-62.
- Kumar, S.S., Kumar, L., Sahai, V., and Gupta, R. (2009). A thiol-activated lipase from *Trichosporon asahii* MSR 54: detergent compatibility and presoak formulation for oil removal from soiled cloth at ambient temperature. *Journal of industrial microbiology & biotechnology* 36, 427-432.
- Labourdenne, S., Cagna, A., Delorme, B., Esposito, G., Verger, R., and Riviere, C. (1997). Oil-drop tensiometer: applications for studying the kinetics of lipase action. *Methods in enzymology* 286, 306-326.

- Laemmli, U.K. (1970).** Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227, 680-685.
- Larsen, M.W., Bornscheuer, U.T., and Hult, K. (2008).** Expression of *Candida antarctica* lipase B in *Pichia pastoris* and various *Escherichia coli* systems. *Protein expression and purification* 62, 90-97.
- Lauersen, K.J., Berger, H., Mussnug, J.H., and Kruse, O. (2013).** Efficient recombinant protein production and secretion from nuclear transgenes in *Chlamydomonas reinhardtii*. *Journal of biotechnology* 167, 101-110.
- Lawyer, F.C., Stoffel, S., Saiki, R.K., Myambo, K., Drummond, R., and Gelfand, D.H. (1989).** Isolation, characterization, and expression in *Escherichia coli* of the DNA polymerase gene from *Thermus aquaticus*. *The Journal of biological chemistry* 264, 6427-6437.
- Lenfant, N., Hotelier, T., Velluet, E., Bourne, Y., Marchot, P., and Chatonnet, A. (2013).** ESTHER, the database of the alpha/beta-hydrolase fold superfamily of proteins: tools to explore diversity of functions. *Nucleic Acids Res* 41, D423-429.
- Li-Beisson, Y., Shorosh, B., Beisson, F., Andersson, M.X., Arondel, V., Bates, P.D., Baud, S., Bird, D., Debono, A., Durrett, T.P., et al. (2010).** Acyl-lipid metabolism. *The Arabidopsis book / American Society of Plant Biologists* 8, e0133.
- Li, M., Ou, X., Yang, X., Guo, D., Qian, X., Xing, L., and Li, M. (2011).** Isolation of a novel C18-Delta 9 polyunsaturated fatty acid specific elongase gene from DHA-producing *Isochrysis galbana* H29 and its use for the reconstitution of the alternative Delta 8 pathway in *Saccharomyces cerevisiae*. *Biotechnology letters* 33, 1823-1830.
- Li, X., Benning, C., and Kuo, M.H. (2012a).** Rapid triacylglycerol turnover in *Chlamydomonas reinhardtii* requires a lipase with broad substrate specificity. *Eukaryotic cell* 11, 1451-1462.
- Li, X., Moellering, E.R., Liu, B., Johnny, C., Fedewa, M., Sears, B.B., Kuo, M.H., and Benning, C. (2012b).** A galactoglycerolipid lipase is required for triacylglycerol accumulation and survival following nitrogen deprivation in *Chlamydomonas reinhardtii*. *The Plant cell* 24, 4670-4686.
- Lie, Ø., and Lambertsen, G. (1986).** Fatty Acid Specificity of *Candida cylindracea* Lipase. *Fette, Seifen, Anstrichmittel* 88, 365-367.
- Lin, K., Simossis, V.A., Taylor, W.R., and Heringa, J. (2005).** A simple and fast secondary structure prediction method using hidden neural networks. *Bioinformatics* 21, 152-159.
- Lin, M., and Oliver, D.J. (2008).** The role of acetyl-coenzyme a synthetase in Arabidopsis. *Plant physiology* 147, 1822-1829.
- Lingane, J.J. (1948).** Automatic Potentiometric Titrations. *Analytical Chemistry* 20, 285-292.
- Liu, W., Jia, B., Zhao, H., Xu, L., and Yan, Y. (2010).** Preparation of a whole-cell biocatalyst of *Aspergillus niger* lipase and its practical properties. *Journal of agricultural and food chemistry* 58, 10426-10430.
- Lombard, S., Helmy, M.E., and Pieroni, G. (2001).** Lipolytic activity of ricin from *Ricinus sanguineus* and *Ricinus communis* on neutral lipids. *The Biochemical journal* 358, 773-781.
- Lopez-Lopez, O., Fucinos, P., Pastrana, L., Rua, M.L., Cerdan, M.E., and Gonzalez-Siso, M.I. (2010).** Heterologous expression of an esterase from *Thermus thermophilus* HB27 in *Saccharomyces cerevisiae*. *Journal of biotechnology* 145, 226-232.
- Lotti, M., Grandori, R., Fusetti, F., Longhi, S., Brocca, S., Tramontane, A., and Alberghina, L. (1993).** Cloning and analysis of *Candida cylindracea* lipase sequences. *Gene* 124, 45-55.
- Lotti, M., Tramontano, A., Longhi, S., Fusetti, F., Brocca, S., Pizzi, E., and Alberghina, L. (1994).** Variability within the *Candida rugosa* lipases family. *Protein engineering* 7, 531-535.
- Lowe, M.E. (1992).** The catalytic site residues and interfacial binding of human pancreatic lipase. *The Journal of biological chemistry* 267, 17069-17073.
- Lowe, M.E. (1997).** Structure and function of pancreatic lipase and colipase. *Annual review of nutrition* 17, 141-158.
- Lowe, M.E. (1999).** Assays for pancreatic triglyceride lipase and colipase. *Methods in molecular biology* 109, 59-70.
- Lubertozzi, D., and Keasling, J.D. (2009).** Developing *Aspergillus* as a host for heterologous expression. *Biotechnology advances* 27, 53-75.

- Lutz, D., Huffer, M., Gerlach, D., and Schreier, P. (1992). Carboxylester-Lipase-Mediated Reactions. In Flavor Precursors (American Chemical Society), pp. 32-45.
- Madzak, C., Gaillardin, C., and Beckerich, J.-M. (2004). Heterologous protein expression and secretion in the non-conventional yeast *Yarrowia lipolytica*: a review. *Journal of biotechnology* 109, 63-81.
- Makrides, S.C. (1996). Strategies for achieving high-level expression of genes in *Escherichia coli*. *Microbiological reviews* 60, 512-538.
- Manderson, D., Dempster, R., and Chisti, Y. (2006). A recombinant vaccine against hydatidosis: production of the antigen in *Escherichia coli*. *Journal of industrial microbiology & biotechnology* 33, 173-182.
- Margreiter, G., Schwanninger, M., Bayer, K., and Obinger, C. (2008). Impact of different cultivation and induction regimes on the structure of cytosolic inclusion bodies of TEM1-beta-lactamase. *Biotechnology journal* 3, 1245-1255.
- Marshall, S.D., Putterill, J.J., Plummer, K.M., and Newcomb, R.D. (2003). The carboxylesterase gene family from *Arabidopsis thaliana*. *Journal of molecular evolution* 57, 487-500.
- Mateos-Diaz, E., Rodriguez, J.A., de Los Angeles Camacho-Ruiz, M., and Mateos-Diaz, J.C. (2012). High-throughput screening method for lipases/esterases. *Methods in molecular biology* 861, 89-100.
- Matos, J.E., Harmon, R.J., and Langlois, B.E. (1995). Lecithinase reaction of *Staphylococcus aureus* strains of different origin on Baird-Parker medium. *Letters in applied microbiology* 21, 334-335.
- Maurich, V., Zacchigna, M., and Pitotti, A. (1991). p-nitrophenyllaurate: a substrate for the high-performance liquid chromatographic determination of lipase activity. *Journal of chromatography* 566, 453-459.
- Mayer, K.M., and Shanklin, J. (2005). A structural model of the plant acyl-acyl carrier protein thioesterase FatB comprises two helix/4-stranded sheet domains, the N-terminal domain containing residues that affect specificity and the C-terminal domain containing catalytic residues. *The Journal of biological chemistry* 280, 3621-3627.
- Mayer, K.M., and Shanklin, J. (2007). Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *BMC plant biology* 7, 1.
- McAleer, W.J., Buynak, E.B., Maigetter, R.Z., Wampler, D.E., Miller, W.J., and Hilleman, M.R. (1984). Human hepatitis B vaccine from recombinant yeast. *Nature* 307, 178-180.
- McKean, S.C., Davies, J.K., and Moore, R.J. (2007). Expression of phospholipase D, the major virulence factor of *Corynebacterium pseudotuberculosis*, is regulated by multiple environmental factors and plays a role in macrophage death. *Microbiology* 153, 2203-2211.
- Meinhardt, F., Stahl, U., and Ebeling, W. (1989). Highly efficient expression of homologous and heterologous genes in *Bacillus megaterium*. *Applied microbiology and biotechnology* 30, 343-350.
- Mijakovic, I. (2010). Protein phosphorylation in bacteria. *Issues*.
- Misset, O., Gerritse, G., Jaeger, K.E., Winkler, U., Colson, C., Schanck, K., Lesuisse, E., Dartois, V., Blaauw, M., Ransac, S., *et al.* (1994). The structure-function relationship of the lipases from *Pseudomonas aeruginosa* and *Bacillus subtilis*. *Protein engineering* 7, 523-529.
- Mnisi, S.M., Louw, M.E., and Theron, J. (2005). Cloning and characterization of a carboxylesterase from *Bacillus coagulans* 81-11. *Current microbiology* 50, 196-201.
- Morrow, K.J., Jr. (2007). Advances in antibody manufacturing using mammalian cells. *Biotechnology annual review* 13, 95-113.
- Mu, H., Kalo, P., Xu, X., and Høy, C.-E. (2000). Chromatographic methods in the monitoring of lipase-catalyzed interesterification. *European Journal of Lipid Science and Technology* 102, 202-211.
- Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. *Cold Spring Harbor symposia on quantitative biology* 51 Pt 1, 263-273.
- Murphy, D.J. (2001). The biogenesis and functions of lipid bodies in animals, plants and microorganisms. *Progress in lipid research* 40, 325-438.
- Muto, S., and Beevers, H. (1974). Lipase Activities in Castor Bean Endosperm during Germination. *Plant physiology* 54, 23-28.

- Nachlas, M.M., and Blackburn, R. (1958). The colorimetric determination of urinary lipase. *The Journal of biological chemistry* 230, 1051-1061.
- Nagao, C., Nagano, N., and Mizuguchi, K. (2014). Prediction of detailed enzyme functions and identification of specificity determining residues by random forests. *PloS one* 9, e84623.
- Neugnot, V., Moulin, G., Dubreucq, E., and Bigey, F. (2002). The lipase/acyltransferase from *Candida parapsilosis*: molecular cloning and characterization of purified recombinant enzymes. *European journal of biochemistry / FEBS* 269, 1734-1745.
- Ntamack, A.G., Karpichev, I.V., Gould, S.J., Small, G.M., and Schulz, H. (2009). Oleate beta-oxidation in yeast involves thioesterase but not Yor180c protein that is not a dienoyl-CoA isomerase. *Biochimica et biophysica acta* 1791, 371-378.
- O'Brien, E.A., Koski, L.B., Zhang, Y., Yang, L., Wang, E., Gray, M.W., Burger, G., and Lang, B.F. (2007). TBestDB: a taxonomically broad database of expressed sequence tags (ESTs). *Nucleic Acids Res* 35, D445-451.
- Oku, H., Koizumi, N., Okumura, T., Kobayashi, T., and Umino, T. (2006). Molecular characterization of lipoprotein lipase, hepatic lipase and pancreatic lipase genes: effects of fasting and refeeding on their gene expression in red sea bream *Pagrus major*. *Comparative biochemistry and physiology Part B, Biochemistry & molecular biology* 145, 168-178.
- Ollis, D.L., Cheah, E., Cygler, M., Dijkstra, B., Frolow, F., Franken, S.M., Harel, M., Remington, S.J., Silman, I., Schrag, J., *et al.* (1992). The alpha/beta hydrolase fold. *Protein engineering* 5, 197-211.
- Ortelli, F., Rossiter, L.C., Vontas, J., Ranson, H., and Hemingway, J. (2003). Heterologous expression of four glutathione transferase genes genetically linked to a major insecticide-resistance locus from the malaria vector *Anopheles gambiae*. *The Biochemical journal* 373, 957-963.
- Ory, R.L., and Altschul, A.M. (1962). A lipid cofactor for the acid lipase of the castor bean. *Biochemical and biophysical research communications* 7, 375-379.
- Ory, R.L., Kiser, J., and Pradel, P.A. (1969). Studies on positional specificity of the castor bean acid lipase. *Lipids* 4, 261-264.
- Padham, A.K., Hopkins, M.T., Wang, T.W., McNamara, L.M., Lo, M., Richardson, L.G., Smith, M.D., Taylor, C.A., and Thompson, J.E. (2007). Characterization of a plastid triacylglycerol lipase from *Arabidopsis*. *Plant physiology* 143, 1372-1384.
- Pahoja, V.M., and Sethar, M.A. (2002). A review of enzymatic properties of lipase in plants, animals and microorganisms. *Pak J Appl Sci* 2, 474-484.
- Palomo, J.M., Fuentes, M., Fernandez-Lorente, G., Mateo, C., Guisan, J.M., and Fernandez-Lafuente, R. (2003). General trend of lipase to self-assemble giving bimolecular aggregates greatly modifies the enzyme functionality. *Biomacromolecules* 4, 1-6.
- Pan, D., Hill, A.P., Kashou, A., Wilson, K.A., and Tan-Wilson, A. (2011). Electrophoretic transfer protein zymography. *Analytical biochemistry* 411, 277-283.
- Panaiotov, I., and Verger, R. (2000). Enzymatic reactions at interfaces: interfacial and temporal organization of enzymatic lipolysis (Adam Baszkin, Willem Norde).
- Pandey, A., Benjamin, S., Soccol, C.R., Nigam, P., Krieger, N., and Soccol, V.T. (1999). The realm of microbial lipases in biotechnology. *Biotechnology and applied biochemistry* 29, 119-131.
- Peabody, D.S. (1989). Translation initiation at non-AUG triplets in mammalian cells. *The Journal of biological chemistry* 264, 5031-5035.
- Pencrac'h, G., and Baratti, J.C. (1997). Activity of *Pseudomonas cepacia* lipase in organic media is greatly enhanced after immobilization on a polypropylene support. *Applied microbiology and biotechnology* 47, 630-635.
- Pereira, S.L., Leonard, A.E., Huang, Y.S., Chuang, L.T., and Mukerji, P. (2004). Identification of two novel microalgal enzymes involved in the conversion of the omega 3-fatty acid, eicosapentaenoic acid, into docosahexaenoic acid. *Biochemical Journal* 384, 357-366.
- Pesaresi, A., and Lamba, D. (2005). Crystallization, X-ray diffraction analysis and phasing of carboxylesterase PA3859 from *Pseudomonas aeruginosa*. *Biochimica et biophysica acta* 1752, 197-201.

- Peti, W., and Page, R. (2007). Strategies to maximize heterologous protein expression in *Escherichia coli* with minimal cost. *Protein expression and purification* 51, 1-10.
- Pistillo, D., Manzi, A., Tino, A., Boyle, P.P., Graziani, F., and Malva, C. (1998). The *Drosophila melanogaster* lipase homologs: a gene family with tissue and developmental specific expression. *Journal of molecular biology* 276, 877-885.
- Pleiss, J., Scheib, H., and Schmid, R.D. (2000). The His gap motif in microbial lipases: a determinant of stereoselectivity toward triacylglycerols and analogs. *Biochimie* 82, 1043-1052.
- Plewniak, F., Bianchetti, L., Brelivet, Y., Carles, A., Chalmel, F., Lecompte, O., Mochel, T., Moulinier, L., Muller, A., Muller, J., *et al.* (2003). PipeAlign: A new toolkit for protein family analysis. *Nucleic Acids Res* 31, 3829-3832.
- Pohnert, G. (2002). Phospholipase A2 activity triggers the wound-activated chemical defense in the diatom *Thalassiosira rotula*. *Plant physiology* 129, 103-111.
- Price, G.M., and Russell, G.B. (1975). Metabolism of beta-[3-H]ecdysone during the larval-pupal stage of the blowfly *Calliphora erythrocephala*. *Biochemical Society transactions* 3, 75-78.
- Pryor, K.D., and Leiting, B. (1997). High-level expression of soluble protein in *Escherichia coli* using a His6-tag and maltose-binding-protein double-affinity fusion system. *Protein expression and purification* 10, 309-319.
- Qi, B., Beaudoin, F., Fraser, T., Stobart, A.K., Napier, J.A., and Lazarus, C.M. (2002). Identification of a cDNA encoding a novel C18-Delta(9) polyunsaturated fatty acid-specific elongating activity from the docosahexaenoic acid (DHA)-producing microalga, *Isochrysis galbana*. *FEBS letters* 510, 159-165.
- Radakovits, R., Jinkerson, R.E., Darzins, A., and Posewitz, M.C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryotic cell* 9, 486-501.
- Radivojac, P., Clark, W.T., Oron, T.R., Schnoes, A.M., Wittkop, T., Sokolov, A., Graim, K., Funk, C., Verspoor, K., Ben-Hur, A., *et al.* (2013). A large-scale evaluation of computational protein function prediction. *Nature methods* 10, 221-227.
- Rahman, R., Leow, T., Salleh, A., and Basri, M. (2007). *Geobacillus zalihae* sp. nov., a thermophilic lipolytic bacterium isolated from palm oil mill effluent in Malaysia. *BMC microbiology* 7, 77.
- Reis, P., Holmberg, K., Watzke, H., Leser, M.E., and Miller, R. (2009). Lipases at interfaces: a review. *Advances in colloid and interface science* 147-148, 237-250.
- Reyes-Duarte, D., Ferrer, M., and Garcia-Arellano, H. (2012). Functional-based screening methods for lipases, esterases, and phospholipases in metagenomic libraries. *Methods in molecular biology* 861, 101-113.
- Riggs, A.D. (1981). Bacterial production of human insulin. *Diabetes care* 4, 64-68.
- Roessler, P.G. (1990). Purification and Characterization of Acetyl-CoA Carboxylase from the Diatom *Cyclotella cryptica*. *Plant physiology* 92, 73-78.
- Rose, K., Rudge, S.A., Frohman, M.A., Morris, A.J., and Engebrecht, J. (1995). Phospholipase D signaling is essential for meiosis. *Proceedings of the National Academy of Sciences of the United States of America* 92, 12151-12155.
- Rudge, S.A., Morris, A.J., and Engebrecht, J. (1998). Relocalization of Phospholipase D Activity Mediates Membrane Formation During Meiosis. *The Journal of cell biology* 140, 81-90.
- Ruiz, C., Falcocchio, S., Pastor, F.I., Saso, L., and Diaz, P. (2007). *Helicobacter pylori* EstV: identification, cloning, and characterization of the first lipase isolated from an epsilon-proteobacterium. *Applied and environmental microbiology* 73, 2423-2431.
- Rupani, B., Kodam, K., Gadre, R., and Najafpour, G.D. (2012). - Lipase-mediated hydrolysis of flax seed oil for selective enrichment of α -linolenic acid. - 114, - 1253.
- Rusch, S.L., and Kendall, D.A. (1995). Protein transport via amino-terminal targeting sequences: common themes in diverse systems (Review). *Molecular membrane biology* 12, 295-307.
- Ruschak, A.M., Rose, J.D., Coughlin, M.P., and Religa, T.L. (2013). Engineered solubility tag for solution NMR of proteins. *Protein science : a publication of the Protein Society*.
- Salovuori, I., Makarow, M., Rauvala, H., Knowles, J., and Kaariainen, L. (1987). Low Molecular Weight High-Mannose Type Glycans in a Secreted Protein of the Filamentous Fungus *Trichoderma reesei*. *Nat Biotech* 5, 152-156.

- Samad, M.Y.A., Razak, C.N.A., Salleh, A.B., Zin Wan Yunus, W.M., Ampon, K., and Basri, M. (1989). A plate assay for primary screening of lipase activity. *Journal of Microbiological Methods* 9, 51-56.
- Sanchez Puerta, M.V., Bachvaroff, T.R., and Delwiche, C.F. (2005). The complete plastid genome sequence of the haptophyte *Emiliania huxleyi*: a comparison to other plastid genomes. *DNA research : an international journal for rapid publication of reports on genes and genomes* 12, 151-156.
- Sandstrom, A.G., Wikmark, Y., Engstrom, K., Nyhlen, J., and Backvall, J.E. (2012). Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library. *Proceedings of the National Academy of Sciences of the United States of America* 109, 78-83.
- Sarda, L., and Desnuelle, P. (1958). [Actions of pancreatic lipase on esters in emulsions]. *Biochimica et biophysica acta* 30, 513-521.
- Satoh, T., and Hosokawa, M. (2006). Structure, function and regulation of carboxylesterases. *Chemico-biological interactions* 162, 195-211.
- Satoh, T., and Hosokawa, M. (2010). Carboxylesterases: structure, function and polymorphism in mammals. *Journal of Pesticide Science* 35, 218-228.
- Saxena, R., Ghosh, P., Gupta, R., Davidson, W.S., Bradoo, S., and Gulati, R. (1999). Microbial lipases: potential biocatalysts for the future industry. *Current Science* 77, 101-115.
- Schein, A.I., Kissinger, J.C., and Ungar, L.H. (2001). Chloroplast transit peptide prediction: a peek inside the black box. *Nucleic Acids Res* 29, E82.
- Schlegel, S., Rujas, E., Ytterberg, A.J., Zubarev, R.A., Luirink, J., and de Gier, J.W. (2013). Optimizing heterologous protein production in the periplasm of *E. coli* by regulating gene expression levels. *Microbial cell factories* 12, 24.
- Schmid, R.D., and Verger, R. (1998). Lipases: Interfacial Enzymes with Attractive Applications. *Angewandte Chemie International Edition* 37, 1608-1633.
- Schue, M., Maurin, D., Dhoub, R., Bakala N'Goma, J.C., Delorme, V., Lambeau, G., Carriere, F., and Canaan, S. (2010). Two cutinase-like proteins secreted by *Mycobacterium tuberculosis* show very different lipolytic activities reflecting their physiological function. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 24, 1893-1903.
- Seitz, E.W. (1974). Industrial application of microbial lipases: a review. *J Am Oil Chem Soc* 51, 12-16.
- Selvy, P.E., Lavieri, R.R., Lindsley, C.W., and Brown, H.A. (2011). Phospholipase D: Enzymology, Functionality, and Chemical Modulation. *Chemical Reviews* 111, 6064-6119.
- Shahi, P., Kumar, I., Sharma, R., Sanger, S., and Jolly, R.S. (2006). Characterization of a novel long-chain acyl-CoA thioesterase from *Alcaligenes faecalis*. *The FEBS journal* 273, 2374-2387.
- Shahidi, F., and Janak Kamil, Y.V.A. (2001). Enzymes from fish and aquatic invertebrates and their application in the food industry. *Trends in Food Science & Technology* 12, 435-464.
- Sharma, R., Chisti, Y., and Banerjee, U.C. (2001). Production, purification, characterization, and applications of lipases. *Biotechnology advances* 19, 627-662.
- Shelley, A.W., Deeth, H.C., and MacRae, I.C. (1987). Comparison of a simple butterfat agar medium with other media used for isolation and enumeration of lipolytic bacteria from dairy products. *The Journal of dairy research* 54, 413-420.
- Shi, B., Wu, W., Wen, J., Shi, Q., and Wu, S. (2010). Cloning and expression of a lipase gene from *Bacillus subtilis* FS1403 in *Escherichia coli*. *Annals of Microbiology* 60, 399-404.
- Shi, T., Yu, A., Li, M., Ou, X., Xing, L., and Li, M. (2012). Identification of a novel C22-4-producing docosahexaenoic acid (DHA) specific polyunsaturated fatty acid desaturase gene from *Isochrysis galbana* and its expression in *Saccharomyces cerevisiae*. *Biotechnology letters* 34, 2265-2274.
- Shockey, J., Chapital, D., Gidda, S., Mason, C., Davis, G., Klasson, K.T., Cao, H., Mullen, R., and Dyer, J. (2011). Expression of a lipid-inducible, self-regulating form of *Yarrowia lipolytica* lipase LIP2 in *Saccharomyces cerevisiae*. *Applied microbiology and biotechnology* 92, 1207-1217.
- Sierra, G. (1957). A simple method for the detection of lipolytic activity of micro-organisms and some observations on the influence of the contact between cells and fatty substrates. *Antonie van Leeuwenhoek* 23, 15-22.

- Singh, S., Kaur, G., Chakraborti, A.K., Jain, R.K., and Banerjee, U.C. (2006). Study of the experimental conditions for the lipase production by a newly isolated strain of *Pseudomonas aeruginosa* for the enantioselective hydrolysis of (+/-)-methyl trans-3(4-methoxyphenyl) glycidate. *Bioprocess and biosystems engineering* 28, 341-348.
- Sinha, J., Plantz, B.A., Inan, M., and Meagher, M.M. (2005). Causes of proteolytic degradation of secreted recombinant proteins produced in methylotrophic yeast *Pichia pastoris*: Case study with recombinant ovine interferon- τ . *Biotechnology and bioengineering* 89, 102-112.
- Sinnett, D., Richer, C., and Baccichet, A. (1998). Rapid zymogram for lipase. *Proc Natl Acad Sci USA* 89, 8794-8797.
- Smichi, N., Fendri, A., Zarai, Z., Bouchaala, E., Chérif, S., Gargouri, Y., and Miled, N. (2012). Lipolytic activity levels and colipase presence in digestive glands of some marine animals. *Fish physiology and biochemistry* 38, 1449-1458.
- Smith, G.M., Rothwell, K., Wood, S.L., Yeaman, S.J., and Bownes, M. (1994). Specificity and localization of lipolytic activity in adult *Drosophila melanogaster*. *The Biochemical journal* 304 (Pt 3), 775-779.
- Snapp, E. (2005). Design and use of fluorescent fusion proteins in cell biology. *Current protocols in cell biology* / editorial board, Juan S Bonifacino [et al] *Chapter 21*, Unit 21 24.
- Sohn, J.H., Kang, H.A., Rao, K.J., Kim, C.H., Choi, E.S., Chung, B.H., and Rhee, S.K. (2001). Current status of the anticoagulant hirudin: its biotechnological production and clinical practice. *Applied microbiology and biotechnology* 57, 606-613.
- Soliman, N.A., Knoll, M., Abdel-Fattah, Y.R., Schmid, R.D., and Lange, S. (2007). Molecular cloning and characterization of thermostable esterase and lipase from *Geobacillus thermoleovorans* YN isolated from desert soil in Egypt. *Process Biochemistry* 42, 1090-1100.
- Song, J.M., An, Y.J., Kang, M.H., Lee, Y.H., and Cha, S.S. (2012). Cultivation at 6-10 degrees C is an effective strategy to overcome the insolubility of recombinant proteins in *Escherichia coli*. *Protein expression and purification* 82, 297-301.
- Song, Y., Choi, M.H., Park, J.N., Kim, M.W., Kim, E.J., Kang, H.A., and Kim, J.Y. (2007). Engineering of the yeast *Yarrowia lipolytica* for the production of glycoproteins lacking the outer-chain mannose residues of N-glycans. *Applied and environmental microbiology* 73, 4446-4454.
- Soni, K.G., Lehner, R., Metalnikov, P., O'Donnell, P., Semache, M., Gao, W., Ashman, K., Pshezhetsky, A.V., and Mitchell, G.A. (2004). Carboxylesterase 3 (EC 3.1.1.1) is a major adipocyte lipase. *The Journal of biological chemistry* 279, 40683-40689.
- Sorensen, J.D., Petersen, E.I., and Wiebe, M.G. (2007). Production of *Fusarium solani* f. sp. pisi cutinase in *Fusarium venenatum* A3/5. *Biotechnology letters* 29, 1227-1232.
- Souleyre, E.J., Marshall, S.D., Oakeshott, J.G., Russell, R.J., Plummer, K.M., and Newcomb, R.D. (2011). Biochemical characterisation of MdCXE1, a carboxylesterase from apple that is expressed during fruit ripening. *Phytochemistry* 72, 564-571.
- Spieß, C., Beil, A., and Ehrmann, M. (1999). A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell* 97, 339-347.
- Spiro, R.G. (2002). Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12, 43R-56R.
- Staubli, F., Jorgensen, T.J., Cazzamali, G., Williamson, M., Lenz, C., Sondergaard, L., Roepstorff, P., and Grimmelikhuijzen, C.J. (2002). Molecular identification of the insect adipokinetic hormone receptors. *Proceedings of the National Academy of Sciences of the United States of America* 99, 3446-3451.
- Staubmann, R., Ncube, I., Gübitz, G.M., Steiner, W., and Read, J.S. (1999). Esterase and lipase activity in *Jatropha curcas* L. Seeds. *Journal of biotechnology* 75, 117-126.
- Stephens, L.L., Shonhai, A., and Blatch, G.L. (2011). Co-expression of the *Plasmodium falciparum* molecular chaperone, PfHsp70, improves the heterologous production of the antimalarial drug target GTP cyclohydrolase I, PfGCHI. *Protein expression and purification* 77, 159-165.
- Sternby, B., Larsson, A., and Borgstrom, B. (1983). Evolutionary studies on pancreatic colipase. *Biochimica et biophysica acta* 750, 340-345.

- Stewart, E.J., Aslund, F., and Beckwith, J. (1998). Disulfide bond formation in the *Escherichia coli* cytoplasm: an in vivo role reversal for the thioredoxins. *The EMBO journal* 17, 5543-5550.
- Stránský, K., Zarevúcka, M., Kejík, Z., Wimmer, Z., Macková, M., and Demnerová, K. (2007). Substrate specificity, regioselectivity and hydrolytic activity of lipases activated from *Geotrichum* sp. *Biochemical Engineering Journal* 34, 209-216.
- Sun, Q.-X., Xue-Ying, L., De-Song, Z., Jiang, L., Xin-Zheng, L., and Bao-Xiu, Q. (2012). Isolation and functional analysis of a $\Delta 5$ desaturase gene from *Isochrysis galbana*. *ACTA HYDROBIOLOGICA SINICA* 36, 412-419.
- Tailor, P.B., Saikia, T.K., Advani, S.H., and Mukhopadhyaya, R. (2004). Activation of HHV-6 in lymphoproliferative disorders: a polymerase chain reaction-based study. *Annals of the New York Academy of Sciences* 1022, 282-285.
- Taketani, M., Shii, M., Ohura, K., Ninomiya, S., and Imai, T. (2007). Carboxylesterase in the liver and small intestine of experimental animals and human. *Life sciences* 81, 924-932.
- Tardif, M., Atteia, A., Specht, M., Cogne, G., Rolland, N., Brugiere, S., Hippler, M., Ferro, M., Bruley, C., Peltier, G., *et al.* (2012). PredAlgo: a new subcellular localization prediction tool dedicated to green algae. *Molecular biology and evolution* 29, 3625-3639.
- Tats, A., Remm, M., and Tenson, T. (2006). Highly expressed proteins have an increased frequency of alanine in the second amino acid position. *BMC genomics* 7, 28.
- Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems. *Applied microbiology and biotechnology* 72, 211-222.
- Thirstrup, K., Verger, R., and Carriere, F. (1994). Evidence for a pancreatic lipase subfamily with new kinetic properties. *Biochemistry* 33, 2748-2756.
- Thompson, P.A., Guo, M.-x., Harrison, P.J., and Whyte, J.N.C. (1992). Effects of variation in temperature on the fatty acid composition of eight species of marine phytoplankton. *Journal of Phycology* 28, 488-497.
- Thomson, C.A., Delaquis, P.J., and Mazza, G. (1999). Detection and measurement of microbial lipase activity: a review. *Critical reviews in food science and nutrition* 39, 165-187.
- Thuren, T., Virtanen, J.A., Somerharju, P.J., and Kinnunen, P.K. (1988). Phospholipase A2 assay using an intramolecularly quenched pyrene-labeled phospholipid analog as a substrate. *Analytical biochemistry* 170, 248-255.
- Tian, W., and Skolnick, J. (2003). How well is enzyme function conserved as a function of pairwise sequence identity? *Journal of molecular biology* 333, 863-882.
- Titball, R.W. (1993). Bacterial phospholipases C. *Microbiological reviews* 57, 347-366.
- Tonon, T., Harvey, D., Larson, T.R., and Graham, I.A. (2003). Identification of a very long chain polyunsaturated fatty acid Delta4-desaturase from the microalga *Pavlova lutheri*. *FEBS letters* 553, 440-444.
- Townsend, A.J., Fields, W.R., Doss, A.J., Clapper, M.L., Doehmer, J., and Morrow, C.S. (1999). Modeling the chemoprotective functions of glutathione S-transferases in cultured cell lines by heterologous expression. *Drug metabolism reviews* 31, 43-69.
- Treichel, H., Oliveira, D., Mazutti, M., Luccio, M., and Oliveira, J.V. (2010). A Review on Microbial Lipases Production. *Food and Bioprocess Technology* 3, 182-196.
- Tripathi, N.K. (2009). High yield production of heterologous proteins with *Escherichia coli*, Vol 59.
- Troeschel, S.C., Thies, S., Link, O., Real, C.I., Knops, K., Wilhelm, S., Rosenau, F., and Jaeger, K.E. (2012). Novel broad host range shuttle vectors for expression in *Escherichia coli*, *Bacillus subtilis* and *Pseudomonas putida*. *Journal of biotechnology* 161, 71-79.
- Tsuchiya, T., Ohta, H., Okawa, K., Iwamatsu, A., Shimada, H., Masuda, T., and Takamiya, K. (1999). Cloning of chlorophyllase, the key enzyme in chlorophyll degradation: finding of a lipase motif and the induction by methyl jasmonate. *Proceedings of the National Academy of Sciences of the United States of America* 96, 15362-15367.
- Tsujisaka, Y., IWAI, M., and TOMINAGA, Y. (1973). Purification, crystallization and some properties of lipase from *Geotrichum candidum* Link. *Agricultural and Biological Chemistry* 37, 1457-1464.

- Tsuya, T., Ferri, S., Fujikawa, M., Yamaoka, H., and Sode, K. (2006). Cloning and functional expression of glucose dehydrogenase complex of *Burkholderia cepacia* in *Escherichia coli*. *Journal of biotechnology* 123, 127-136.
- Turk, B., Turk, V., and Turk, D. (1997). Structural and functional aspects of papain-like cysteine proteinases and their protein inhibitors. *Biological chemistry* 378, 141-150.
- Upadhyay, A.K., Murmu, A., Singh, A., and Panda, A.K. (2012). Kinetics of inclusion body formation and its correlation with the characteristics of protein aggregates in *Escherichia coli*. *PloS one* 7, e33951.
- Upton, C., and Buckley, J.T. (1995). A new family of lipolytic enzymes? *Trends in biochemical sciences* 20, 178-179.
- Vakhlu, J., and Kour, A. (2006). Yeast lipases: enzyme purification, biochemical properties and gene cloning.
- Valero, F. (2012). Heterologous expression systems for lipases: a review. *Methods in molecular biology* 861, 161-178.
- Verdoes, J.C., Punt, P.J., and Hondel, C.A.M.J.J. (1995). Molecular genetic strain improvement for the overproduction of fungal proteins by Filamentous fungi. *Applied microbiology and biotechnology* 43, 195-205.
- Verger, R. (1976). Interfacial enzyme kinetics of lipolysis. *Annual review of biophysics and bioengineering* 5, 77-117.
- Verger, R. (1980). Enzyme kinetics of lipolysis. *Methods in enzymology* 64, 340-392.
- Verger, R. (1984). Lipases. *Lipases*.
- Villeneuve, P., and Foglia, T. (1997). Lipase specificities: Potential application in lipid bioconversions. *Inform* 8, 640-650.
- Vincent, F., Charnock, S.J., Verschueren, K.H., Turkenburg, J.P., Scott, D.J., Offen, W.A., Roberts, S., Pell, G., Gilbert, H.J., Davies, G.J., *et al.* (2003). Multifunctional xylooligosaccharide/cephalosporin C deacetylase revealed by the hexameric structure of the *Bacillus subtilis* enzyme at 1.9Å resolution. *Journal of molecular biology* 330, 593-606.
- von Tigerstrom, R.G., and Stelmaschuk, S. (1989). The use of Tween 20 in a sensitive turbidimetric assay of lipolytic enzymes. *Canadian journal of microbiology* 35, 511-514.
- Wahler, D., Badalassi, F., Crotti, P., and Reymond, J.-L. (2001). Enzyme Fingerprints by Fluorogenic and Chromogenic Substrate Arrays. *Angewandte Chemie International Edition* 40, 4457-4460.
- Wang, D., Xu, Y., and Shan, T. (2008). Effects of oils and oil-related substrates on the synthetic activity of membrane-bound lipase from *Rhizopus chinensis* and optimization of the lipase fermentation media. *Biochemical Engineering Journal* 41, 30-37.
- Wang, G., Ryu, S., and Wang, X. (2012). Plant phospholipases: an overview. *Methods in molecular biology* 861, 123-137.
- Wang, X. (2000). Multiple forms of phospholipase D in plants: the gene family, catalytic and regulatory properties, and cellular functions. *Progress in lipid research* 39, 109-149.
- Wang, X. (2001). Plant Phospholipases. *Annual review of plant physiology and plant molecular biology* 52, 211-231.
- Warden, C.H., Davis, R.C., Yoon, M.Y., Hui, D.Y., Svenson, K., Xia, Y.R., Diep, A., He, K.Y., and Lusi, A.J. (1993). Chromosomal localization of lipolytic enzymes in the mouse: pancreatic lipase, colipase, hormone-sensitive lipase, hepatic lipase, and carboxyl ester lipase. *Journal of lipid research* 34, 1451-1455.
- Weslake, R.J., Thomson, L.W., Tenaschuk, D., and Mackenzie, S.L. (1989). Properties of Solubilized Microsomal Lipase from Germinating *Brassica napus*. *Plant physiology* 91, 1303-1307.
- West, N.P., Chow, F.M., Randall, E.J., Wu, J., Chen, J., Ribeiro, J.M., and Britton, W.J. (2009). Cutinase-like proteins of *Mycobacterium tuberculosis*: characterization of their variable enzymatic functions and active site identification. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* 23, 1694-1704.
- Westers, L., Westers, H., and Quax, W.J. (2004). *Bacillus subtilis* as cell factory for pharmaceutical proteins: a biotechnological approach to optimize the host organism. *Biochimica et biophysica acta* 1694, 299-310.

- Wilson, P.A., Gardner, S.D., Lambie, N.M., Commans, S.A., and Crowther, D.J. (2006). Characterization of the human patatin-like phospholipase family. *Journal of lipid research* 47, 1940-1949.
- Winkler, U.K., and Stuckmann, M. (1979). Glycogen, hyaluronate, and some other polysaccharides greatly enhance the formation of exolipase by *Serratia marcescens*. *Journal of bacteriology* 138, 663-670.
- Wong, H., and Schotz, M.C. (2002). The lipase gene family. *Journal of lipid research* 43, 993-999.
- Wong, S.L., Kawamura, F., and Doi, R.H. (1986). Use of the *Bacillus subtilis* subtilisin signal peptide for efficient secretion of TEM beta-lactamase during growth. *Journal of bacteriology* 168, 1005-1009.
- Wu, X., Jornvall, H., Berndt, K.D., and Oppermann, U. (2004). Codon optimization reveals critical factors for high level expression of two rare codon genes in *Escherichia coli*: RNA stability and secondary structure but not tRNA abundance. *Biochemical and biophysical research communications* 313, 89-96.
- Xu, Y., Yasin, A., Tang, R., Scharer, J.M., Moo-Young, M., and Chou, C.P. (2008). Heterologous expression of lipase in *Escherichia coli* is limited by folding and disulfide bond formation. *Applied microbiology and biotechnology* 81, 79-87.
- Yamada, J., Suga, K., Furihata, T., Kitahara, M., Watanabe, T., Hosokawa, M., Satoh, T., and Suga, T. (1998). cDNA cloning and genomic organization of peroxisome proliferator-inducible long-chain acyl-CoA hydrolase from rat liver cytosol. *Biochemical and biophysical research communications* 248, 608-612.
- Yamane, T., Suzuki, T., Sahashi, Y., Vikersveen, L., and Hoshino, T. (1992). Production of n-3 polyunsaturated fatty acid-enriched fish oil by lipase-catalyzed acidolysis without solvent. *J Am Oil Chem Soc* 69, 1104-1107.
- Yatsu, L.Y., and Jacks, T.J. (1972). Spherosome membranes: half unit-membranes. *Plant physiology* 49, 937-943.
- Yeaman, S.J. (2004). Hormone-sensitive lipase--new roles for an old enzyme. *The Biochemical journal* 379, 11-22.
- Yildir, C., Onsan, Z.I., and Kirdar, B. (1998). Optimization of starting time and period of induction and inducer concentration in the production of the restriction enzyme *EcoRI* from recombinant *Escherichia coli* 294. *Turk J Chem* 22, 221-226.
- Yin, J., Li, G., Ren, X., and Herrler, G. (2007). Select what you need: a comparative evaluation of the advantages and limitations of frequently used expression systems for foreign genes. *Journal of biotechnology* 127, 335-347.
- Yoon, K., Han, D., Li, Y., Sommerfeld, M., and Hu, Q. (2012). Phospholipid:diacylglycerol acyltransferase is a multifunctional enzyme involved in membrane lipid turnover and degradation while synthesizing triacylglycerol in the unicellular green microalga *Chlamydomonas reinhardtii*. *The Plant cell* 24, 3708-3724.
- Young, J.C., Agashe, V.R., Siegers, K., and Hartl, F.U. (2004). Pathways of chaperone-mediated protein folding in the cytosol. *Nature reviews Molecular cell biology* 5, 781-791.
- Yu, J., Hu, S., Wang, J., Wong, G.K., Li, S., Liu, B., Deng, Y., Dai, L., Zhou, Y., Zhang, X., *et al.* (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). *Science* 296, 79-92.
- Yu, S., Zheng, B., Zhao, X., and Feng, Y. (2010). Gene cloning and characterization of a novel thermophilic esterase from *Fervidobacterium nodosum* Rt17-B1. *Acta Biochim Biophys Sin (Shanghai)* 42, 288-295.
- Yuan, L., Voelker, T.A., and Hawkins, D.J. (1995). Modification of the substrate specificity of an acyl-acyl carrier protein thioesterase by protein engineering. *Proceedings of the National Academy of Sciences of the United States of America* 92, 10639-10643.
- Zhang, J., Liu, J., Zhou, J., Ren, Y., Dai, X., and Xiang, H. (2003). Thermostable esterase from *Thermoanaerobacter tengcongensis*: high-level expression, purification and characterization. *Biotechnology letters* 25, 1463-1467.
- Zhang, X., Li, M., Agrawal, A., and San, K.Y. (2011). Efficient free fatty acid production in *Escherichia coli* using plant acyl-ACP thioesterases. *Metabolic engineering* 13, 713-722.
- Zhang, Y. (2009). Lipase Diversity in Glacier Soil Based on Analysis of Metagenomic DNA Fragments and Cell Culture. *Journal of Microbiology and Biotechnology* 19, 888-897.

- Zheng, Y., Li, L., Liu, Q., Qin, W., Yang, J., Cao, Y., Jiang, X., Zhao, G., and Xian, M. (2012). Boosting the free fatty acid synthesis of *Escherichia coli* by expression of a cytosolic *Acinetobacter baylyi* thioesterase. *Biotechnology for biofuels* 5, 76.
- Zhou, C., Yang, Y., and Jong, A.Y. (1990). Mini-prep in ten minutes. *BioTechniques* 8, 172-173.

Thèse de Doctorat

Vincent KERVIEL

Clonage et caractérisation de deux gènes codant des enzymes lipolytiques de la microalgue *Isochrysis galbana*

Cloning and characterization of lipolytic enzymes from two isolated genes of *Isochrysis galbana*

Résumé

Les enzymes lipolytiques sont des ester hydrolases impliquées dans le métabolisme lipidique. Leurs caractéristiques se sont révélées être des atouts dans de nombreuses applications industrielles. Chez les microalgues, l'isolement et la caractérisation de ces enzymes d'un point de vue structural et fonctionnel restent des domaines de recherche peu explorés à ce jour.

Certaines espèces bénéficient pourtant de contenus en lipides intéressants, source de matière première pour les industries de l'agroalimentaire ou de l'énergie. Par exemple, l'acide docosahexaénoïque (DHA), un acide gras polyinsaturé de la série des oméga 3, est reconnu pour ses propriétés en santé humaine. Parmi de nombreuses espèces, *Isochrysis galbana*, une microalgue unicellulaire appartenant à la classe des *Prymnesiophycées* est considérée comme une source possible de DHA. La présence d'acides gras libres a été montrée par l'analyse des lipides, suggérant la présence d'enzymes lipolytiques potentiellement intéressantes pour leur sélectivité et leur spécificité de substrat.

L'analyse d'une banque de marqueurs de séquences exprimées a permis l'identification de séquences susceptibles de coder des enzymes lipolytiques. Les ARN messagers ont été extraits et les ADN complémentaires ont été clonés.

Ce travail de thèse présente l'analyse et le clonage de deux gènes codant une ester hydrolase putative et une thioestérase putative, issues de la microalgue *Isochrysis galbana*.

Les deux séquences codent des protéines de poids moléculaires de 35,41 kDa et de 42,31 kDa. Elles montrent 30 à 40 % d'identité et de similarité avec des hydrolases α/β notamment des carboxylestérases de différents organismes. Les séquences protéiques ont permis l'identification du pentapeptide consensus Gly-X-Ser-X-Gly caractéristique des enzymes lipolytiques et les acides aminés Ser/Asp/His de la triade catalytique.

Les deux séquences codantes ont été clonées et exprimées dans la levure *Saccharomyces cerevisiae* et la bactérie *Escherichia coli*. Le clonage dans *E. coli* a permis d'identifier à la taille attendue une protéine par Western blot. En présence de cette protéine, la composition en acides gras des lipides de la bactérie a été modifiée. L'analyse CPG a notamment montré une augmentation des proportions en acides gras C16 :1 et C18 :1 par rapport au témoin. Ce résultat permet de caractériser l'activité thioestérase pour *IgTeCe*.

Mots clés

Ester hydrolase, *Isochrysis galbana*, Expression de protéines, Acides gras

Abstract

Lipolytic enzymes present in all known species play a key role in lipid metabolism and are involved in several industrial processes. They catalyse lipid hydrolysis and synthesis. Actually and particularly in microalgae, isolation and characterization of this type of enzyme remains an unexplored research area.

The potential of the lipidic content of microalgae in food industry or energy field requires specific lipolytic enzymes. Docosahexaenoic acid (DHA), an $\omega 3$ poly unsaturated fatty acid ($\omega 3$ PUFA) is well known for its beneficial effects on human health. Among many species, *Isochrysis galbana*, a unicellular marine microalga belonging to the *Prymnesiophyceae* class, is considered as a potential alternative source of DHA.

Lipid analysis of *I. galbana* shows free fatty acids and suggests the presence of lipolytic enzymes with potential interesting selectivities and substrate specificities. Analysis of incomplete expressed sequence tag (EST) listed in the EST bank of *Isochrysis galbana*, identified incomplete genes that encode lipolytic enzymes. Messenger RNAs were extracted, characterized and cloned.

This work describes the analysis and cloning of two genes encoding a putative ester hydrolase and a putative thioesterase in marine microalgae *Isochrysis galbana*. Sequences encode two proteins with predicted molecular weights of approximately 35,41 kDa and 42,31 kDa. Slight similarity and identity (from 30 to 40 %) were observed between the gene sequence and various α/β fold hydrolase found in diverse phyla (including carboxylesterase).

Sequences also included the consensus Gly-X-Ser-X-Gly and the catalytic triad Ser/Asp/His. To characterize the predicted enzymatic functions, an experimental procedure was introduced: coding sequences were cloned into expression vectors and expressed in *Saccharomyces cerevisiae* and in *Escherichia coli*.

Western blot identification of recombinant enzyme shows a convenient protein production in bacteria. Furthermore, the expression of the protein in *E. coli* shifted the fatty acid composition predominantly towards C16:1 and C18:1 fatty acids. The enzyme called *IgTeCe* showed a thioesterase activity.

Key Words

Ester hydrolase, *Isochrysis galbana*, Protein expression, Fatty acids